

농어가경제조사 교체가구 추정기법
개발 연구

2004년 10월

한국조사연구학회

농어가경제조사 교체가구 추정기법
개발 연구

2004년 10월

한국조사연구학회

제 출 문

통계청장 귀하

본 보고서를 『농어가경제조사 교체가구 추정기법 개발』
의 연구결과보고서로 제출합니다.

2004년 10월

책임연구원 김규성 (서울시립대학교 통계학과)

공동연구원 이기재 (한국방송통신대학교 정보통계학과)

머 리 말

본 보고서는 통계청으로부터 연구용역을 의뢰받아 수행한 “농어가경제조사 교체가구 추정기법 개발” 최종보고서이다.

통계청에서는 우리나라 농어가경제 및 농어업 경영실태를 파악하기 위하여 농어가경제조사를 실시하고 있다. 농어가경제조사 표본에 포함된 농어가는 일정기간 동안 표본의 임무를 맡게 되는데, 농어업 환경의 변화와 표본 농어가의 사정에 의하여 표본에서 탈퇴하는 농어가가 발생하고 있고, 또한 표본 농어가의 사정에 의하여 혹은 표본 교체 과정에서 응답을 받지 못하는 경우가 발생하고 있다. 이러한 경우는 표본교체와 무응답 처리를 하여 조사의 연속성을 유지하여야 한다. 표본교체와 무응답은 사전에 방지하는 것이 최우선이겠지만, 현실적으로 발생한 표본교체와 무응답에 대해서는 사후적으로 적절하게 처리하여 조사의 신뢰도를 유지하는 일이 중요하다.

본 보고서는 농어가경제조사에서 발생하는 표본교체와 무응답을 통계적으로 처리하는 방법에 관한 연구 결과이다. 표본교체 및 무응답 처리를 위하여 농어가경제조사에서 발생하는 표본교체 및 무응답 특성을 고찰하였고, 이를 바탕으로 표본을 교체하는 방법과 무응답을 처리하는 방법을 제안하였다. 무응답 처리의 효율을 높이기 위하여 대체군을 형성하였고, 가중하트를 수행한 후 대체 후 추정량에 대한 분산 추정 방법을 제안하였다. 이를 구현하는 SAS 프로그램을 부록에 수록하였다. 본 연구에서 제안하는 표본 교체 방법과 무응답 처리 기법이 농어가경제조사에 활용되어 향후 농어가경제조사의 질적인 수준을 높이는데 일조할 것으로 기대한다.

끝으로 본 연구를 수행하는데 적극적인 협조를 아끼지 않은 통계청 허남거 과장, 김진 사무관께 깊은 감사를 드리며, 연구기간 동안 조언을 아끼지 않은 농수산통계과 직원 여러분께도 감사를 드리는 바이다. 그리고 자료정리 및 분산추정기법 모의 실험에 많은 시간과 노력을 아끼지 않은 서울시립대학교 통계학과 대학원생 곽수연 양에게 심심한 사의를 표한다.

2004년 10월

연구책임자

서울시립대학교 교수 김규성

차 례

1. 서 론	1
1.1 개요	1
1.2 연구의 내용과 방법.....	3
2. 현행 표본설계 고찰	5
2.1 표본설계	5
2.2 추정방법	8
2.3 일계부와 원부데이터 결합.....	10
3. 농어가표본의 교체 및 무응답 처리	12
3.1 표본농어가의 교체 특성	12
3.2 표본농어가의 무응답 특성	13
3.3 교체 및 무응답 처리 후 추정	15
4. 무응답 대체군 형성	20
4.1 대체군 형성 방법	20
4.2 농가경제조사를 위한 대체군	22
4.3 어가경제조사를 위한 대체군	44
5. 가중핫덱대체 및 분산추정	60
5.1 가중핫덱대체	60
5.2 잭나이프 분산추정	61
5.3 모의실험	63
5.4 잭나이프 분산추정 : 다중대체군.....	66
6. 결론	69
참고문헌	71
<부록> 잭나이프 분산추정량 계산 알고리즘 (SAS 프로그램).....	72

1. 서 론

1.1 개요

농어가경제조사는 우리나라 농어가경제 및 농어업 경영실태를 파악하여 농수산 정책 및 농수산 문제 연구의 기초자료를 제공하기 위하여 실시되고 있다. 농가경제 조사는 1953년부터, 어가경제조사는 1963년부터 실시해온 계속조사로써, 농어가에서 발생하는 날마다의 수입, 지출을 조사하고 또한 농어가의 재산이나 부채의 변동 사항을 조사한다. 현행 농어가경제조사는 날마다의 수입, 지출 상황을 기록하는 일계부조사와, 연중의 재산 변동을 기록하는 원부조사를 병행하여 실시하고 있는데, 일계부 조사는 매일 기록하여 월 단위로 집계가 이루어지고 있으며, 원부 조사는 연초와 연말에 조사하여 그 차이를 집계하고 있다. 연간 통계는 일계부 집계와 원부 집계가 결합되어 작성된다.

표본농어가는 표본설계를 통하여 전국의 농어가를 대표할 수 있도록 선정되며, 표본으로 선정된 농어가는 5년간 표본농어가로서의 임무를 맡게 된다. 표본설계는 농어업총조사 자료를 바탕으로 작성되므로 5년에 한번씩 재설계가 이루어질 수밖에 없다. 따라서 정상적인 상황에서 표본농어가는 5년 후에 자연 교체가 이루어지게 된다. 그러나 우리나라 농어촌의 변화를 반영하여 표본 농어가에도 개별적인 변화가 있을 수 있으며, 또한 표본 농어가로서의 임무에 대한 피로감 때문에 표본에서 탈퇴를 하는 경우도 발생할 수 있다. 즉, 표본설계 시에 농어가로 선정될 때에는 농어가였으나 전업 등으로 농어가의 기준에서 벗어나는 일반 가구로 전환되거나, 이사 등으로 타 지역으로 이동하는 경우, 그리고 응답의 피로감으로 인하여 응답을 거절하는 경우가 발생하게 된다. 이러한 경우는 조사의 연속성 및 조사 결과의 신뢰도를 유지하기 위하여 탈퇴하는 농어가와 농어가 형태 및 소득 구조가 가장 유사한 농어가로 교체하고 있다.

최근의 농어가경제조사 현황을 보면, 1998년 설계시의 농가 표본수는 3,140가구, 어가 표본수는 1,360가구 였으며, 2002년 재설계시 농가표본수는 3,200가구, 어가 표본수는 1,175가구였다. 이 중에서 교체가 일어난 농어가 수는 아래의 <표 1.1>과 같다.

<표 1.1> 농어가경제조사 교체가구수

연도 \ 조사	농가경제조사	어가경제조사
1999 년	106 (3.4%)	77 (5.7%)
2000 년	100 (3.2%)	68 (5.0%)
2001 년	99 (3.2%)	80 (4.9%)
2002 년	59 (1.9%)	30 (2.2%)
2003 년	149 (4.7%)	96 (8.2%)

최근 5년간의 교체율은 표본농가의 경우 3.3%, 표본어가의 경우는 5.2%에 달하고 있다. 그리고 표본 개편 첫해(2003년의 경우)에 표본 교체가 가장 많이 일어나고 있으며, 해가 지날수록 교체의 비율은 서서히 감소하는 것으로 나타나고 있다. 그러나 작년의 경우 교체 비율이 재작년의 경우와는 현격하게 늘어난 수치를 보이고 있는데, 이는 조사환경이 갈수록 어려워지고 있음을 반영하는 것으로 풀이할 수 있고 더 나아가 교체율이 과거보다는 더 증가할 것이라고 예상할 수 있다.

한 시점 조사의 경우 유사한 속성을 갖는 농어가로 표본 교체를 하면 표본수가 유지되기 때문에 조사 결과의 신뢰도는 크게 변하지 않으므로 전반적인 조사 과정에 큰 문제는 발생하지 않는다. 그러나 계속조사에서 동일한 농어를 계속 조사하는 경우는 상황이 다르다. 즉, 농어가 경제지표의 경향을 유지하려면 동일한 가구의 조사결과를 계속 집계해야 하는데, 교체로 인하여 조사가구가 바뀌면 표본교체로 인한 효과가 경제지표에 영향을 미칠 수 있는 것이다. 또한 교체를 하는 과정에서 새로운 농어를 표본에 편입하는 시기가 늦어져 일정기간 동안 공백이 생기기도 한다. 이 경우 조사를 실시하지 못한 기간은 표본 자료가 없으므로 시계열 유지를 위해서는 사후적인 처리가 뒤따라야 한다.

표본조사에서 모집단 및 표본의 상황 변동으로 인한 표본교체 및 무응답 발생은 불가피한 것으로 보인다. 원천적으로 표본 교체 및 무응답 발생을 방지할 수는 없기 때문에 표본교체율 및 무응답 발생률을 최소한으로 하는 것이 최선일 것이다. 그러나 일단 이루어진 교체 및 무응답에 대해서는 사후적인 통계 처리를 가미하는 것이 일반적인 추세이고, 이에 대한 여러 가지 기법들이 최근에 많이 개발되고 있다. 본 연구는 후자에 관련한 것이다. 즉, 교체 및 무응답이 발생한 상황에서 교체 및 무응답의 특성을 고찰하고, 교체 및 무응답으로 인한 효과를 파악하여 추정과정

에 반영함으로써 월평균 혹은 연평균 추정치 및 추정치에 대한 신뢰도 추정치가 타당하도록 하는 방법을 개발하는 것이다.

1.2 연구 내용과 방법

농어가경제조사에서 오차가 포함되는 과정은 크게 4부분으로 나누어 볼 수 있다.

첫째는 표본농어를 선정하는 과정인데, 표본조사는 모집단 전체를 조사하는 것이 아니고 모집단의 일부인 표본만을 조사하기 때문에 전체가 아닌 일부를 조사함으로써 발생하는 표본오차는 불가피하다. 표본선정은 모집단을 대표할 수 있는 표본을 선정하여 대표성을 확보하는 일이 무엇보다 중요하다. 농어가경제조사는 모두 층화 후 집락을 추출하고 표본 집락에서 다시 농어를 추출하는 층화 2단집락 추출법을 사용하였다.

두 번째로, 일단 표본으로 선정된 농어가는 5년 동안 표본의 임무를 수행하게 되는데, 이 기간에 표본으로 선정된 농어가에서 교체가 발생하고 교체가 발생하는 과정에서 응답을 받지 못하는 경우가 생긴다. 이러한 경우, 교체율을 최소한으로 하고 무응답율을 최소한으로 하는 것이 바람직하다. 그리고 불가피한 교체 및 무응답에 대해서는 사후적인 처리가 뒤따라야 한다. 표본 교체 및 무응답이 발생하면 이로 인한 추가적인 오차가 포함될 소지가 있으므로 이에 대한 적절한 대처가 필요하다.

세 번째로는, 데이터를 수집하는 과정에서 오차가 개입될 수 있다. 일계부 및 원부 작성 시에 정확한 응답을 받도록 해야 하며, 응답 오차를 최소한으로 줄이는 노력이 필요하다.

마지막으로, 수집한 데이터를 이용하여 집계를 한 후, 월별 혹은 연간 추정치를 만드는데 오차가 개입될 수 있다. 즉, 농어가경제조사에서 표본은 층화 2단집락추출법에 의하여 선정이 되었으므로, 설계에 기반한 추정공식을 이용하여 추계를 해야 모평균 등에 대한 비편향 추정치를 얻을 수 있다. 설계와 다른 추정 공식을 이용하면 편향 추정치를 얻을 수 있으며, 편향 추정치는 추계하고자 하는 대상을 과소 혹은 과대로 계산하게 되어 신뢰도가 떨어질 수 있다. 또한 농어가경제조사에서는 교체 및 무응답이 발생하므로, 이로 인한 효과가 추정 과정에 반영되어야 한다. 교체 및 무응답 효과가 추정과정에 반영되지 않으면 추계결과에 편향이 개입될 수 있다.

본 연구는 교체 및 무응답 처리 방법, 그리고 이로 인한 효과를 추정과정에 반영하는 방법에 연관되어 있으므로, 위의 과정에서 두 번째 과정과 네 번째 과정과 밀접한 관련이 있다. 그러나 교체 및 무응답은 모두 현행 표본설계에 의하여 선정된 표본을 바탕으로 하므로 현행 표본설계와 무관하게 이루어질 수는 없으며, 현행 표본설계를 밑바탕으로 하여 교체 효과 및 대체 효과 그리고 이를 반영한 추정기법을 개발해야 한다. 따라서 포괄적으로 본 연구는 위의 네 과정 중 세 번째 과정을 제외한 전체 과정과 관련이 있다고 할 수 있다.

본 보고서의 구성은 다음과 같다. 제 2장에서는 현행 표본설계에 대한 간단한 검토가 이루어진다. 그리고 일계부 데이터와 원부 데이터의 관계를 검토하여 일계부 데이터와 원부데이터의 효과를 비교 분석한다. 제 3장에서는 표본 교체와 무응답으로 인한 효과를 검토한다. 표본 농어가의 교체 특성과 무응답 특성을 살펴보고 교체 및 무응답으로 인한 효과를 추정에 반영하는 방법을 모색해 본다. 제 4장은 무응답 대체와 관련한 장이다. 무응답 대체를 하기 위해서는 무응답 대체군 형성이 필수적인데, 4장에서는 농가경제조사와 어가경제조사에 적합한 대체군 형성을 제안할 것이다. 제 5장에서는 가중핫텍 대체 후 분산추정을 하는 방법에 관련한 장이다. 가중핫텍 대체 절차를 소개하고, 가중핫텍 대체 후 분산추정법을 설명한다. 마지막으로 제 6장에서는 이제까지의 연구 내용을 간략히 요약한다.

2. 현행 표본설계 고찰

2.1 표본설계

2.1.1 농가경제조사 표본설계

농가경제조사 표본설계는 각 도를 부차모집단으로 하여 각 도에서 층화 2단집락 추출법을 사용하였다. 이때 1차 추출단위로는 농업조사구나 부락이 선정되었으며, 2차 추출단위로는 농가가 사용되었다. 또한 층내에서는 동일한 추출확률을 사용하여 동일층 내에서의 가중치는 동일하게 하였다. 결과적으로 각 도별로 10개의 층을 만들었으며, 320개의 조사구(12개 주산지 포함)가 선정하였고, 각 조사구에서 10 농가씩 총 3,200 농가가 표본으로 선정하였다. 층화 및 표본배정은 <표 2.1>에 요약되어 있고 <표 2.2>에 지역별 표본조사구수 현황이 나타나 있다.

<표 2.1> 농가경제조사 표본설계의 층화 기준

구 분	층 화 기 준
논벼지구1	전업 또는 1종 겸업농가이면서 논벼 농가의 비율이 70%이상인 지구
논벼지구2	전업 또는 1종 겸업농가이면서 논벼 농가의 비율이 50%-70%인 지구
과수지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 과수 농가가 최대인 지구(10% 이상)
채소지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 채소 농가가 최대인 지구(10% 이상)
특작지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 특작 농가가 최대인 지구(10% 이상)
화훼지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 화훼 농가가 최대인 지구(10% 이상)
전작지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 전작 농가가 최대인 지구(10% 이상)
축산지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 축산 농가가 최대인 지구(10% 이상)
기타지구	위의 영농특성 층이나 2종겸업 층에 속하지 않는 지구
2종겸업	2종 겸업농가의 비율이 50%이상인 지구

<표 2.2> 각 지역별 층별 표본조사구 수 현황

	논벼1	논벼2	과수	채소	특작	화훼	전작	축산	기타	2종	합계
경기	4	6	3	6	1	4	2	7	2	3	38
강원	4	6	1	9	1	0	7	2	1	3	34
충북	3	6	7	8(1)	4	0	1	2	1	2	34
충남	7	8	4	7	3	0	1	4	2	2	38
전북	12(1)	9	2	6	2	0	1	2	1	1	36
전남	11(1)	7	4	9(2)	1	0	3	2	1	3	41
경북	5	7(2)	10	10(2)	3	0	1	3	1	2	42
경남	5	7	4	11(2)	1	2	1	4	1	3	39
제주	0	0	10	3	0	0	2(1)	0	0	3	18
전국	51	56	45	69	16	6	19	26	10	22	320

2.1.2 어가경제조사 표본설계

어가경제조사 표본설계의 특징은 층화 2단 집락추출을 사용한 점과, 층내에서는 동일한 추출확률을 사용한 점이다. 따라서 층내에서 동일한 추출확률을 사용했기 때문에 동일 층 내에서는 동일한 가중치가 부여된다. 최종적인 층의 수는 20개이며, 276조사구에서 1,175 어가가 표본으로 선정되었다. <표 2.3>에 층화 및 조사구수, 표본 어가수 등이 요약되어 있다.

<표 2.3> 어가경제조사 : 층별 조사구 크기, 표본크기, 가중값

	층번호	모집단 어가수	모집단 조사구수	표본 조사구수	표본어가수 (전,1종)	가중값	표본어가수 (2종)	가중값
전남 전북	01	1347	54	14	56	24.0536	1	263.244
	02	9895	324	35	140	70.6786	4	263.244
	03	2295	101	6	24	96.6250	2	263.244
	04	3186	127	12	48	66.3750	2	263.244
	05	4757	197	21	84	56.6310	8	263.244
경남 경북 충남	11	812	36	10	40	20.3000	0	0
	12	2,821	94	21	84	33.5833	5	263.244
	13	3,866	130	22	88	43.9318	4	263.244
	14	1,940	68	11	44	44.0909	4	263.244
	15	3,702	157	14	56	66.1071	4	263.244
	16	3,895	148	13	52	74.9038	5	263.244
	17	3,806	171	8	32	118.9375	3	263.244
부산 인천 경기	21	988	36	10	40	24.7000	3	263.244
	22	3,376	102	14	56	60.2857	5	263.244
	23	1,654	61	4	16	103.3750	0	0
울산 강원 제주	31	1,572	45	15	60	26.2000	8	263.244
	32	1,638	64	12	48	34.1250	5	263.244
	33	5,083	222	19	76	66.8816	8	263.244
전업 1종 규모 이하	50	2,299	988	15	18	65.8667	15	263.244
					15	87.8222		
					12	109.7778		
2종 접업	90	22,639						
합 계		81,571		276	1,089		86	

2.2 추정방법

농어가경제조사 표본설계에는 집계 항목에 대한 월별 모평균을 추정하는 공식이 제안되어 있다. 제안된 추정공식은 교체 및 무응답이 없는 경우에 사용할 수 있도록 고안된 것이며, 또한 한 시점에서 모평균을 추정하는 것이 초점이 맞추어져 있다.

그런데 현실적으로는 조사과정에서 교체 및 무응답이 발생하고, 연간 통계는 일계부의 월별 집계 자료와 원부의 연간 자료를 결합하여 작성하므로 이에 대한 추가적인 추정과정이 필요하다. 일계부와 원부의 결합 과정은 다음절에서 살펴보기로 하고, 교체 및 무응답 효과는 다음 장에서 고찰할 것이다. 본 절에서는 현행 표본설계에서 제시하는 모평균 추정공식과 표준오차 추정공식, 그리고 현행표본설계의 효과를 알아보기 위하여 설계효과와 급내상관계수를 구하여 본다.

2.2.1 농가경제조사

모평균 추정량

농가경제조사 표본설계(2002년)가 제안하는 항목별 모평균 추정량은 가중값을 이용한 가중평균의 형태이다.

- $w_{...} = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{k=1}^{m_{hi}} w_{hik}$: 표본내 가중값의 합
- w_{hik} : h 번째 층의 i 번째 조사구의 k 번째 표본가구에 대한 가중값
- y_{hik} : h 번째 층의 i 번째 조사구의 k 번째 표본가구에 대한 조사값
- $\bar{y}_C = \frac{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hik} y_{hik}}{w_{...}}$: 모평균 추정량

따라서 모평균 추정치를 계산하기 위해서는 항목별로 표본농가에서의 조사값 y_{hik} 및 표본농가에 부여된 가중값 w_{hik} 을 알고 있어야 한다.

상대표준오차 추정량

모평균 추정량의 분산추정량은 선형화 방법에 의한 분산추정량에 각 층에서 구한 표본추출률 $f_h = n_h/N_h$ 를 추가한 추정공식으로 추정한다. 이 공식은 통계패키지 SAS에서도 구현이 가능한 공식이므로 SAS를 이용하여 손쉽게 추정치를 구할 수 있는 장점이 있다.

- $v(\bar{y}_C) = \sum_{h=1}^H \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi} - \bar{e}_{h.})^2$: 분산추정량
- $e_{hi} = \left(\sum_k w_{hik} (y_{hik} - \bar{y}_C) \right) / w_{...}$, $\bar{e}_{h.} = \left(\sum_{i=1}^{n_h} e_{hi} \right) / n_h$
- $se(\bar{y}_C) = \sqrt{v(\bar{y}_C)}$: 표준오차 추정량
- $cv(\bar{y}_C) = se(\bar{y}_C) / \bar{y}_C$: 상대표준오차 추정량

2.2.2 여가경제조사

모평균 추정량

여가경제조사 표본설계(2002년)가 제안하는 항목별 모평균 추정량은 다음과 같다

- $w_{...} = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{k=1}^{m_{hi}} w_{hik}$: 표본내 가중값의 합
- w_{hik} : h 번째 층의 i 번째 조사구의 k 번째 표본가구에 대한 가중값
- y_{hik} : h 번째 층의 i 번째 조사구의 k 번째 표본가구에 대한 조사값
- $\bar{y}_C = \frac{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{k=1}^{m_{hi}} w_{hik} y_{hik}}{w_{...}}$: 모평균 추정량 (가중평균)

따라서 모평균 추정치를 계산하기 위해서는 항목별로 표본여가에서의 조사값 y_{hik} 및 표본여가에 부여된 가중값 w_{hik} 이 필요하다.

상대표준오차 추정량

표본설계가 제안하는 상대표준오차 추정량은 아래와 같다.

- $v(\bar{y}_C) = \sum_{h=1}^H \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi} - \bar{e}_{h.})^2$: 분산추정량
- $e_{hi} = \left(\sum_{j=1}^k w_{hik} (y_{hik} - \bar{y}_C) \right) / w_{...}$, $\bar{e}_{h.} = \left(\sum_{i=1}^{n_h} e_{hi} \right) / n_h$
- $se(\bar{y}_C) = \sqrt{v(\bar{y}_C)}$: 표준오차 추정량
- $cv(\bar{y}_C) = se(\bar{y}_C) / \bar{y}_C$: 상대표준오차 추정량

2.3 일계부와 원부 데이터 결합

농어가경제조사는 일계부 및 원부의 데이터를 결합하여 연평균 추정치를 작성하고 있다. 따라서 일계부 데이터와 원부데이터가 어떻게 결합되며, 일계부 및 원부 데이터가 연간 추정치에 어떤 영향을 주는지 고찰할 필요가 있다. 농어가경제조사에서 원부항목이 일계부항목과 결합되는 형태는 감가상각액, 연초평가액과 연말평가액의 차이, 그리고 증식액이다.

(i) 감가상각액 :

감가상각액이 관여되는 항목은 건물, 기계기구, 무형자산의 농어업용, 겸업용, 가계용, 기타용 감가상각액이다.

(ii) 연초 및 연말평가액 차이 :

연초 및 연말 평가액의 차이가 연간 추계치에 반영되는 항목은 미처분 농축수산물 및 미사용자재 항목이다.

(iii) 증식액 : 증식액은 대동식물 항목에 관련되는데, 다음과 같이 계산된다.

$$\begin{aligned} \text{증식액} &= (\text{연말평가액} + \text{연중 감소평가액(대동식물 판매액)}) \\ &\quad - (\text{연초평가액} + \text{연중 증가평가액(대동식물 구입액)}) \end{aligned}$$

아래의 <표 2.5>에 일계부만 사용한 결과와 일계부 및 원부를 사용한 결과의 비

교가 있다. 비교에 사용한 표본수는 1,155어가(2003년 12월)이다. 이 결과에서 주목할 점은 일부 변수에서 두 모평균이 큰 차이를 보인다는 점이다. 특히, 어가소득의 경우, 일계부 평균은 2,652만원이나 일계부와 원부를 결합할 경우 연평균 소득은 2,921만원에 이른다. 이때 상대편향은 -9.2%이다. 즉, 원부 데이터를 사용하지 않으면 소득이 9.2% 하향 추정된다는 뜻이다. 또한 상대표준오차는 일계부만 사용할 경우 3.2%이며, 원부 데이터가 추가되면 5.1%가 되어 상대편향은 무려 -36.5%에 이른다. 원부데이터의 영향이 매우 큼을 나타내주는 수치이다.

<표 2.5> 일계부 결과와 일계부 및 원부 결과 비교 (단위 : 10,000)

항목명	평균		상대표준오차	
	일계부	일계부+원부	일계부	일계부+원부
어가소득	2652.08	2921.42	0.0326	0.0514
어업소득	1292.96	1295.59	0.0497	0.0497
겸업소득	447.30	714.02	0.0960	0.1770
사업외소득	454.86	454.86	0.0667	0.0667
이전소득	84.51	84.51	0.0826	0.0826
비경상소득	372.43	372.43	0.0981	0.0981
가계비	1990.11	1990.11	0.0236	0.0236
어업수입	2693.61	2695.83	0.0583	0.0583
어로수입	1578.37	1580.59	0.0576	0.0575
양식수입	1041.56	1041.56	0.1386	0.1386
어업잡수입	73.67	73.67	0.1988	0.1988
겸업수입	966.89	1233.64	0.1084	0.1303
농업수입	294.92	561.67	0.1067	0.2134

3. 농어가표본의 교체 및 무응답 처리

3.1 표본농어가의 교체 특성

농가경제조사의 2003년 표본농가는 3,200가구이며, 이중 3,043(95.1%)의 농가에서 는 교체가 일어나지 않았으며, 148 농가(4.6%)에서 1회 이상 교체가 발생하였다. 148 농가 중 표본 교체가 2회 이상 발생한 가구는 모두 9가구이며, 2회 7가구, 3회 2가구이다. 어가경제조사의 2003년 표본어가는 1,175가구이며, 이중 1,079(91.8%)어 가에서는 교체가 일어나지 않았으며, 96어가(8.2%)에서 1회 이상 교체가 발생하였 다. 이중 교체 사유가 확인된 어가는 92어가이다. 92어가 중 표본 교체가 2회 이상 발생한 가구는 모두 4가구이며, 2회 3가구, 4회 1가구이다. 교체가 2회 발생한 3가 구 중 2가구는 연속 교체로 인하여 표본데이터는 수집되지 않았다.

<표 3.1> 농어가경제조사의 교체 패턴

교체회수	농가경제조사	어가경제조사
무교체	3,043	1,079
1회 교체	139	94
2회 교체	7	1
3회 교체	2	1
기타	9	
	3,200	1,175

농가경제조사의 경우 표본으로 선정된 농가가 표본에서 이탈하는 사유로 농가 환경의 변화(비농가, 단독가구, 전출, 장기 출타)가 75%에 이른다. 이러한 변화는 모집 단 변화를 표본이 반영하는 것이므로 변화가 시작된 시점에서 교체해 주어야 한다. 농가경제조사의 경우에 조사불능과 조사불응의 비율은 25%로 나타나고 있다. 어가 경제조사의 경우는 표본으로 선정된 농어가가 표본에서 이탈하는 사유는 어가 환경 의 변화(비어가, 단독어가, 전출, 장기 출타)가 80%에 이른다. 또한 조사불능과 조사 불응의 비율은 20%정도 되는데, 이 경우는 가능한 원래 선정된 어가가 표본으로 남

도록 현지조사원의 노력이 요구되는 부분이다.

<표 3.2> 농어가표본의 교체 사유

교체사유	표본 농가	표본 어가
비농어가	44 (28.0%)	40(39.2%)
전출	25 (15.9%)	21(20.6%)
단독가구	45 (28.7%)	17(16.7%)
조사불능	23 (14.6%)	12(11.8%)
조사불응	15 (9.6%)	8(7.8%)
장기출타	5 (3.2%)	4(3.9%)
합계	157 (100%)	102(100%)

3.2 표본농어가의 무응답 특성

농가경제조사(2003년)

전체 3,200 표본농가 중 무응답이 발생한 농가는 27농가이다. 무응답이 발생한 전체 27가구는 모두 표본농가를 교체하는 과정에서 무응답이 발생하였다. 즉, 대부분의 무응답은 표본농가를 교체하는 과정에서 표본에서 이탈하는 가구와 새로이 표본에 포함되는 가구를 섭외하는 과정에서 발생함을 알 수 있다.

<표 3.3> 농가경제조사의 교체 및 무응답 패턴 (2003년)

교체회수	완전응답	무응답		전체
		중간	연결	
무교체	3,043			3,043
1회 교체	116		23	139
2회 교체	4		3	7
3회 교체	1		1	2
기타	9			9
	3,173		27	3,200

무응답이 지속되는 개월 수를 보면, 26농가에 대해서 1달이며, 2개월이 1가구로

나타났다. 즉, 표본농가의 이탈이 발생하면, 거의 대부분 1달 이내에 새로운 농가로 교체가 되어 확정됨을 보여준다.

<표 3.4> 무응답 지속 개월 수

무응답 지속 개월	1개월	2개월	3개월	4개월	합계
무응답 농가수	26	1	-	-	27

어가경제조사(2003년)

전체 1,175 표본어가 중 무응답이 발생한 어가는 20어가이다. 이 중 1가구(어가 번호 381102)만 계속되는 조사 중간(8월)에 무응답이 발생 했을 뿐, 나머지 19가구는 모두 표본어를 교체하는 과정에서 무응답이 발생하였다. 즉, 대부분의 무응답은 표본어를 교체하는 과정에서 표본에서 이탈하는 어가와 새로이 표본에 포함되는 어가를 섭외하는 과정에서 발생하고 있다.

<표 3.5> 어가경제조사의 교체 및 무응답 패턴(2003년)

교체 회수	완전응답	무응답		전체
		중간	연결	
무교체	1,078		1	1,079
1회 교체	75	1	18	94
2회 교체	1			1
3회 교체	1			1
	1,155	1	19	1,175

무응답이 지속되는 개월 수를 보면, 대부분은 1달이며, 2개월이 3가구, 4개월이 2어가로 나타났다. 즉, 표본어의 이탈이 발생하면, 대부분 1달 이내에 새로운 어가로 교체가 되지만, 25% 정도는 2달 이상이 걸려 새로운 어가가 표본으로 확정됨을 보여준다.

<표 3.6> 무응답 지속 개월 수

무응답 지속 개월	1개월	2개월	3개월	4개월	합계
무응답 어가수	14	3	0	2	19

3.3 교체 및 무응답 처리 후 추정

농어가경제조사에서 표본교체는 추정 효율의 관점이 아니라 표본 농어가 관리의 관점에서 교체가 이루어지고 있다. 그리고 위에서 살펴보았듯이 표본 농어가의 교체는 불가피한 경우가 대부분이다. 또한 무응답률은 높지 않으며(농가경제조사 : 27/3,200=0.8%, 어가경제조사:20/1175=1.7%), 대부분의 무응답은 표본 농어가가 교체되는 과정에서 발생하고 있다. 표본의 대표성을 유지하기 위해서는 가능한 교체나 무응답이 발생하지 않는 것이 바람직하지만, 교체와 무응답이 불가피하게 발생하는 상황에서는 사후적으로 교체와 무응답을 적절하게 처리하는 것이 요구된다.

<표 3.7> 일계부와 원부의 데이터 구조

표본 번호	일계부												원부	
	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	연초	연말
1	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	z1	z2
2	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	z1	z2
3	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	z1	z2
.
1079	x1	x2	x3	x4	?	y6	y7	y8	y9	y10	y11	y12	z1	u2
1080	x1	x2	x3	x4	x5	x6	x7	x8	y9	y10	y11	y12	z1	u2
1081	x1	x2	x3	x4	?	y6	y7	y8	y9	y10	y11	y12	z1	u2
.
n	x1	x2	x3	?	y5	y6	y7	y8	y9	y10	y11	y12	z1	u2

주) x1-x12 : 교체 전 일계부 값, y1-y12 : 교체 후 일계부 값

z1, z2 : 교체 전 원부 값, u2 : 교체 후 원부 값, ? : 무응답

농어가경제조사에서 교체와 무응답이 발생한 상황을 데이터 관점에서 보면 위의 <표 3.7>로 설명할 수 있다. 일계부의 경우는 조사월 사이에 나타나는 무응답을 적절하게 처리해야 하며, 원부의 경우는 교체가 일어난 표본가구에서 연초평가액, 연말평가액, 연중 증감액을 사후적으로 적절하게 처리해야 한다.

3.3.1 무응답 처리

무응답을 사후에 통계적으로 처리하는 방법은 무응답이 단위무응답인지 항목무응답인지에 따라 다르게 처리하는 것이 보통이다. 통상적으로 단위무응답은 가중치 조정을 통하여 처리하며, 항목무응답인 경우는 대체를 통하여 처리하는 것이 일반적이다.

농어가경제조사의 무응답 패턴은 기존의 분류에 의하면 모두가 단위무응답에 속한다. 따라서 통상적인 접근방식으로 보면 가중치 조정을 하는 것이 타당해 보인다. 그러나 농어가경제조사의 속성을 들여다보면 가중치 조정보다는 대체를 하는 것이 더 합리적인데 그 이유는 다음과 같다.

(i) 일계부 조사는 12개월 계속되기 때문에 12개월 중 한 두 달에서 발생한 무응답은 시계열 과정에서 항목 무응답으로 간주할 수 있다.

(ii) 더 근본적인 이유는 추정과정에 있다. 농어가경제조사에는 수십여 개의 항목이 있고, 소분류 항목이 더해져서 중분류 항목을 만들고, 중분류 항목이 더해져서 대분류 항목을 만든다. 또한 일부 항목은 일계부 데이터와 원부 데이터가 결합되어 집계가 된다. 따라서 월별로 무응답 횟수가 다르므로, 월별로 가중치 조정을 하게 되면 12개의 계산 알고리즘이 있어야 하고, 연간 추정에는 또 하나의 계산 알고리즘이 있어야 한다. 현실적으로 계산의 부담이 크다.

이러한 이유로 농어가경제조사에서는 가중치 조정보다는 무응답 대체가 더 합리적인 방법이라고 할 수 있다. 따라서 본 연구에서는 무응답 대체에 관한 방법만을 다루기로 한다.

무응답은 속성에 따라 (i) 무시 가능한 무응답과 (ii) 무시하기 어려운 무응답으로 구분하는데, 무시가 가능한 무응답은 무응답이 발생한 메커니즘이 조사항목과 연관관계가 없는 경우이다. 반대로 무시하기 어려운 무응답은 응답 메커니즘이 조사항목

과 연관성이 있는 경우이다. 예를들어 농어가경제조사에서 주로 발생하는 무응답의 원인으로 농어가의 비농어가화, 단독농어가화, 진출, 조사불능 등은 무시가능한 무응답으로 분류할 수 있다. 반대로 자연재해 등으로 인해 발생한 무응답은 무시하기 어려운 무응답으로 분류할 수 있다. 무시가능한 무응답인 경우는 가중치 조정이나 무응답 대체를 통하여 처리가 가능한 반면, 무시하기 어려운 무응답은 가중치조정이나 대체와는 다른 조치가 필요하다.

자연재해 등으로 인한 무응답은 영역 추정(domain estimation)을 하는 것이 하나의 방법이다. (영역추정에 관한 내용은 Sarndal, et al. (1992), p. 69을 참조). 자연재해로 인해 피해를 입은 지역을 하나의 관심영역으로 분류하여 이 영역에 대한 추정은 다른 지역과 다르게 독립적으로 수행을 하는 것이다. 이 지역은 모두 재해를 입었기 때문에 표본가구가 입은 피해 정도로 이 영역의 피해를 추정하는 것이 합리적이다. 이러한 영역 추정을 할 때 제기될 수 있는 추가적인 사항은 다음과 같다. 첫째, 관심 영역을 정해야 한다. 관심 영역에 대한 범위가 분명하게 정해져야 추계가 가능하기 때문이다. 둘째, 경우에 따라 관심영역에 속하는 표본의 수가 적을 수 있다. 특히 관심 영역이 좁을 때에는 이에 속하는 표본의 수가 적을 가능성이 크다. 이 때는 임시적으로 관심영역 추계만을 위한 추가 표본을 선정하여 추계치의 최소 신뢰도를 확보하는 것이 하나의 방법이 될 수 있다.

표본대체 방법에는 보조변수 이용이 가능한 경우와 가능하지 않는 경우에 구분할 수 있다. 관심변수와 밀접한 관련이 있는 보조변수(예를 들어 관심변수와 상관관계가 높은 변수)를 사전에 알고 있으면 보조변수를 이용하려 비대체나 회귀대체를 하는 것이 효과적일 수 있다. 만일 이용 가능한 보조변수를 확보하기 어려운 경우는 평균대체, 핫덱 대체, 혹은 최근방대체 등이 유력한 방법인데, 평균대체는 대체후 표본분포의 왜곡이 심하다는 단점 때문에 권장되지 않는 방법이며, 최근방대체는 유사한 농어를 대체할 수는 장점이 있지만, 대체 후 표준오차 계산이 불가능하다는 단점 때문에 공식 통계에서는 사용에 어려움이 있다. 반면 대체후 표본분포를 대체전 분포와 유사하게 유지시켜주는 핫덱 대체가 농어가경제조사에서는 가장 권장할 만한 대체 방법이다. 대체방법 및 대체후 추정에 관해서는 5장에서 자세히 설명하기로 한다.

3.3.2 표본 교체 처리

일계부만을 이용하는 경우 교체효과는 항목 추계에 별다른 영향을 미치지 못한다. 즉, 교체전 가구나 교체후 가구를 모두 확률표본으로 간주하면 교체유무에 관계없이 월평균 추계를 하고 이를 합산하여 연평균추계를 하면 된다. 문제는 일계부와 원부가 결합하여 추계되는 항목에 있다. 원부 항목에서는 연초, 연말평가액과 연중 증감액이 추정에 반영하므로, 표본농어가가 교체되면 이러한 수치를 직접적으로 구하기 어려워진다. 이제 교체 농어를 처리하는 방안을 고찰해보자.

교체 농어를 처리하는 방법은 크게 세 가지로 나누어 볼 수 있다.

- (i) 교체된 농어를 이용하지 않는 방법
- (ii) 교체전, 후의 농어를 모두 이용하는 방법
- (iii) 교체후 농어를 기준으로 교체 이전의 상황을 추정하는 방법

교체 농어를 추정에 이용하지 않으면 추정의 문제는 발생하지 않는다. 그러나 농가경제조사의 경우 교체가구는 4.6%에 이르며 어가경제조사의 경우는 8.2%에 이르기 때문에 교체 농어를 추정에 반영하지 않는 것은 상당수의 표본을 조사가 불완전하다는 이유로 사용하지 않는 결과가 되어, 예산은 물론 추정의 효율성에서도 손실을 초래할 우려가 있다. 두 번째로 교체전, 후 가구의 조사값을 동일한 가구의 조사값인 것처럼 간주하는 경우, 교체전 가구의 수치와 교체후 가구의 수치의 차이가 크면 효율성은 저하될 수 있다. 세 번째 방법은 전자의 두 방법 보다는 데이터를 최대한으로 활용하는 방법이므로 효율성은 증가하지만, 현실적으로 교체전 가구의 수치를 어떻게 추정할 것인가 하는 문제가 대두된다.

교체 농어가 처리에 문제가 되는 부분은 교체 시점을 기준으로 볼 때, 연초 시점의 평가액과 연초 및 교체 시점 사이에 발생한 증감액을 파악하는 일이다. 앞의 2.3절에서 살펴본 바와 같이 연초평가액과 증감액 추정에 관여되는 항목은 건물, 기계기구, 무형자산의 감가상각액, 미처분농축산물 및 미사용자재의 연초, 연말 평가액 차이, 그리고 대동식물의 연초, 연말 평가액 차이 및 연중 증식액이다. 두 가지 방법을 고려하자.

(i) 교체시점에서 연초평가액 및 증감액을 추정하는 방법

(ii) 교체시점을 연초로 간주하는 방법

첫 번째 방법의 경우, 교체 시점에서 연초의 원부 사항을 면접 조사하고, 교체월 이전 상황에 대해서도 조사한다. 이 방법은 연초 평가액과 연초부터 교체월 사이의 변동액을 추정하기 때문에 교체전 상황이 추계에 반영되는 장점이 있는 반면, 현실적으로 연초 수치를 제대로 추정할 수 있는지 여부에 문제가 관건이다. 감가상각액 추정은 가능할 것으로 보이는 반면, 미처분농축산물, 미사용자재, 대동식물의 연초 평가액, 증식액에 대해서는 응답자의 기억에 의존해야 하는 단점이 있다.

두 번째로, 교체 시점을 연초 시점으로 간주하여 원부 사항을 조사하면 원부의 세부 사항을 정확하게 조사할 수 있는 장점이 있는 반면, 교체전 상황을 추정에 반영하지 못한다는 단점이 있다. 이 방법 역시 감가상각액은 연초평가액 추정이 가능할 것이므로, 추정이 안되는 항목은 미처분농수산물과 대동식물의 연초평가액과 교체시점까지의 증식액이다.

이 두 방법은 모두 장단점이 있기 때문에 어느 한 방법이 우수하다고 보기는 어렵다. 두 방법 중 하나를 고르는 하나의 선택기준은 연초부터 교체시점 사이에 발생한 미처분농축산물, 미사용자재 및 대동식물의 변동액의 크기이다. 즉, 변동액의 크기가 크지 않으면 교체 시점을 연초시점으로 간주해도 추정에 큰 오차는 수반되지 않을 것이다. 그러나 만일 그 차이가 크다면 교체시점을 연초로 간주하는 것 보다는 추정을 통하여 그 차이를 보완하는 것이 필요하다.

교체후 가구의 가구주로부터 교체전 상황을 기억에 의존하여 미처분농축산물, 미사용자재 및 대동식물 변동액의 크기를 조사하는 것이 분명하지 않다면, 교체전 가구의 일계부에서 나타나는 미처분농축산물의 변동액 및 대동식물의 증감액을 조사하여 교체후 가구의 수치로 이용하는 것도 하나의 대안이 된다. 왜냐하면 일계부는 교체와 관계없이 연속적으로 사용되며 원부의 증감 사항이 일계부와 연동되어 계산되므로, 불분명한 기억에 의존하는 것 보다는 기록된 수치를 사용하는 것이 불확실성이 줄어들기 때문이다.

4. 무응답 대체군 형성

4.1 대체군 형성 방법

가중햇택법을 이용해서 무응답을 대체하기 위해서는 우선 대체군(Imputation class)을 형성하는 것이 필수적이다. 대체군을 형성하는 이유는 무응답으로 인한 편향을 대체를 통하여 최소로 줄이기 위해서이다. 대체군을 만드는 방법은 보조정보를 이용하여 (i) 사후 증화를 하거나 (ii) 무응답이 발생한 가구와 비슷한 가구를 묶어 대체군을 형성하는 방법이 있다. 또는 로지스틱회귀분석을 통하여 응답률을 추정 후, 비슷한 응답률을 갖는 가구를 대체군으로 묶는 방법 등이 있다. 그런데 2003년도 농어가경제조사 결과를 보면 무응답 또는 표본교체 비율이 4.6%-8%로 전체적으로 작게 나타나고 있어서, 응답률이 비슷하도록 대체군을 묶는 방법은 현실적으로 적용하기 어렵다. 본 절에서는 대체군을 형성하는 방법에 관하여 고찰한다.

농어가경제조사는 일계부를 통해서 월 단위로 데이터가 정리되고 있지만 최종적인 통계는 연 단위로 집계하여 발표하고 있다. 본 연구에서 농어가경제조사의 무응답 대체군은 월 단위 데이터와 연 단위 데이터를 종합적으로 분석하여 설정하였다.

햇택대체를 위한 무응답 대체군을 설정할 때 기본원칙은 주요 관심변수들이 동질적이도록 구분하여 대체군을 구성하는 것이다. 이 과정은 증화나 사후증화의 과정과 유사하다고 할 수 있다.

농어가경제조사 주요 관심변수는 소득이나 수입과 같은 연속형 변수이다. 본 연구에서는 2003년도 농어가경제조사 자료에 대해서 SAS Enterprise Miner에서 제공하고 있는 의사결정나무모형(Decision Tree Model)을 이용하여 대체군을 구분하여 결정하는 방법을 사용하였다. 각 가구의 농가소득(어가소득)과 농업소득(어업소득)을 목표변수(target variable)로 농어의 특성을 나타내는 나머지 변수들을 설명변수(input variable)로 하였다.

본 연구에서 적용된 무응답 대체군 설정 과정은 다음과 같다.

1단계 : 관심변수(target variable)로 사용할 변수 선택

농가경제조사에서 대체군 설정에 사용된 주요 관심변수는 농가소득, 농업소득이

다. 어가경제조사도 농가경제조사와 마찬가지로 주요 관심변수는 어가소득, 어업소득으로 한다. 부차적으로 겸업소득과 사업외소득을 고려한다.

2단계 : 관심변수에 영향을 주는 주요 설명변수 선택

주요 관심변수에 영향을 주는 주요 설명변수로 영농형태, 전·겸업, 지역, 가구원수, 가구주 연령 등을 고려한다. 어가경제조사의 주요 설명변수로는 어업형태, 전·겸업, 지역, 선박톤수, 어장면적 등을 고려한다. 이들 변수들이 주요 관심변수에 어떤 영향을 미치는가를 살펴보기 위해서 회귀분석을 실시한다.

3단계 : 대체군 설정 과정에서 사용될 설명변수의 우선순위 결정

회귀분석 결과와 전반적인 영농형태(어업형태), 전·겸업, 지역별 표본 농어가 분포를 확인하여 대체군 설정과 관련하여 사용할 변수들의 우선순위를 결정한다. 농가경제조사에 대해서는 영농형태와 전·겸업 구분을 우선적으로 고려하고, 지역, 경지면적, 농업종사자 수, 경영주 연령 등을 부차적으로 고려한다. 어가경제조사에 대해서는 어업형태와 전·겸업 구분을 우선적으로 고려하고 지역, 선박 톤수 또는 어장 면적을 부차적으로 고려한다.

4단계 : 의사결정 나무모형을 이용하여 대체군 구성

무응답 대체군은 어느 정도 이상의 표본을 포함하고 있어야 하고, 대체군 내에서 관심변수들은 동질적이어야 한다. 그러나 단순히 의사결정 나무모형(decision tree model)을 적용해서 대체군을 구성하면 실제적으로 활용할 수 없거나 대체군 내의 표본이 너무 적게 배정되는 사례가 빈번하게 발생한다. 따라서 농가경제조사의 경우에는 영농형태, 전·겸업, 지역 구분을 고려해서 일차적으로 대체군을 구성하고, 이들 대체군 중에서 표본이 충분히 큰 경우에 대해서 경지면적을 설명변수로 하는 의사결정 나무모형을 적용하여 대체군을 세분하는 작업을 진행한다. 어가경제조사의 경우도 농가경제조사와 유사하게 어업형태, 전·겸업 구분, 지역, 선박 톤수나 양식장 면적 등을 고려해서 대체군을 구성한다.

5단계 : 대체군 구성의 타당성 확인을 위한 검토 작업

농어가경제조사의 주요 변수에 대해서 대체군별 평균과 표준편차 추정값을 계산하여 만들어진 대체군의 타당성을 확인한다.

농어가경제조사를 위한 무응답 대체 방법으로 가중랜덤핫덱 방법, 최근거리 대체 방법을 사용할 수 있을 것이다. 농어가경제조사에서 결측값의 비중은 상당히 낮은 편에 속한다. 따라서 본 연구에서는 추정량의 분산계산이 비교적 간편한 가중랜덤 핫덱 방법을 이용해서 결측값을 대체하고자 한다. 아울러 결측값을 대체할 때는 대체군 내의 자료값 중에서 극단값을 제거하고 대체하는 것이 바람직할 것이다.

4.2 농가경제조사를 위한 대체군

농가경제조사의 대체군 설정과 관련하여 고려한 주요 관심변수는 농가소득, 농업 소득, 겸업소득, 사업외 소득이고, 이들 변수에 영향을 줄 것으로 예상되는 설명변수로 영농형태, 전·겸업, 지역(도), 경지면적, 종사자 수, 연령 등을 고려하였다.

다음 <표 4.1>은 2003년도 농가경제조사에 대해서 전국의 영농형태 및 전·겸업별 표본크기 현황이다. 영농형태와 전·겸업 구분에 따라서 대단한 편차가 있음을 알 수 있다. 대부분의 표본농가는 논벼(전업), 논벼(1종겸업), 과수(전업), 채소(전업), 축산(전업), 2종겸업 등으로 구분됨을 알 수 있다. 나머지 영농형태는 그 수가 상대적으로 대단히 적다.

한편, <표 4.2>는 각 도의 영농형태별 표본크기 현황을 정리한 것이다. 지역적으로 살펴보면 논벼, 채소, 축산, 2종겸업 등은 지역적으로 비교적 고르게 분포하고 있지만 나머지 영농형태는 표본농가의 수가 적기 때문에 지역적으로 세분된 대체군을 형성할 수 없다.

각 도의 영농형태 및 전·겸업별 표본크기 현황을 분석한 결과에 의하면 각 도별로 충분한 표본농가가 있는 논벼(전업), 채소(전업), 2종 겸업농 등에 대해서는 각 도별로 이들 농가들을 각각 묶어서 각 대체군으로 구성할 수 있는 정도의 표본수이다. 다만, 제주도의 논벼(전업)는 경남의 논벼(전업) 대체군과 병합하여 구성한다. 통계분석 결과에 의하면 논벼(전업), 채소(전업), 2종겸업의 층 내에서 지역간의 소득 차이는 없는 것으로 나타났지만, 각 도에 충분한 표본이 배분되어 있기 때문에 지역간의 유사성을 고려하여 무응답 대체군을 구성하는 것이 바람직하다.

<표 4.1> 전국의 영농형태 및 전·겸업별 표본크기 현황

	논벼	과수	채소	특작	화훼	전작	축산	기타	2종겸업	전체
전업	703	178	479	56	24	66	214	15	-	1735
1종겸업	264	55	114	27	5	20	64	5	-	554
2종겸업	-	-	-	-	-	-	-	-	754	754
전체	967	233	593	83	29	86	278	20	754	3043

<표 4.2> 각 도의 영농형태별 표본크기 현황

	논벼	과수	채소	특작	화훼	전작	축산	기타	2종겸업	전체
경기	98	15	41	7	15	5	33	4	134	352
강원	91	7	58	15	2	27	31	2	90	323
충북	83	24	73	15	0	17	41	1	71	325
충남	134	21	66	14	0	2	28	0	99	364
전북	183	8	38	8	0	2	25	0	73	337
전남	161	29	82	4	6	7	36	8	65	398
경북	117	71	98	17	0	2	33	2	59	399
경남	98	15	103	3	2	7	48	2	95	373
제주	2	43	34	0	4	17	3	1	68	172
전체	967	233	593	83	29	86	278	20	754	3043

<표 4.3> 영농형태별, 전·겸업별 농가별 평균 농업종사자 수(단위: 명)

구분	논벼	과수	채소	특작	화훼	전작	축산	기타	2종겸업	전체
전업	2.2	2.29	2.3	2.37	2.14	2.24	2.17	2.27	-	2.23
1종겸업	2.3	2.36	2.44	2.34	2.27	2.26	2.6	3.25	-	2.37
2종겸업	-	-	-	-	-	-	-	-	2.42	2.42
전체	2.21	2.30	2.31	2.36	2.15	2.24	2.23	2.40	2.42	2.29

<표 4.4> 영농형태별, 전·겸업별 농업경영주의 평균 연령(단위: 세)

구분	논벼	과수	채소	특작	화훼	전작	축산	기타	2종겸업	전체
전업	65.00	61.73	58.89	57.04	49.37	63.40	60.12	62.67	-	62.20
1종겸업	59.90	57.00	56.34	52.88	56.95	57.27	54.92	54.72	-	58.02
2종겸업	-	-	-	-	-	-	-	-	56.88	56.88
전체	64.21	60.91	58.64	56.12	50.29	62.53	59.38	61.58	56.88	60.48

<표 4.3>은 영농형태별, 전·겸업별 농가별 평균 농업종사자 수 현황을 정리한 것이다. 가구당 농업종사자 수는 영농형태 구분에 따라서 뚜렷하게 차이 나지는 않았다. 대체적인 경향을 보면 전업농가의 농업종사자 수가 겸업농가에 비해서 다소 작은 것으로 나타났다.

<표 4.4>는 영농형태별, 전·겸업별 농업경영주의 평균 연령 현황이다. 상대적으로 전업농가의 경영주 연령이 높은 것으로 나타났고, 영농형태별 구분에 따라서 살펴보면 화훼, 특작, 2종 겸업 농가의 경영주 연령이 낮게 나타났다.

<표 4.5> 영농형태별, 전·겸업별 가구당 평균 경지면적 현황(단위: 평)

구분		논벼	과수	채소	특작	화훼	전작	축산	기타	2종겸업	전체
논면적	전업	4795.5	1362.0	2164.2	2271.9	871.4	1040.2	2362.2	1139.3	-	3302.2
	1종겸업	6959.1	1539.3	2446.2	1819.7	316.8	1399.7	2957.0	1295.3	-	4668.9
	2종겸업	-	-	-	-	-	-	-	-	1824.8	1824.8
	전체	5130.9	1392.8	2192	2172.7	804	1091.3	2448.6	1160.7	1824.8	3093.9
밭면적	전업	1195.8	3731.4	2485.2	5075.9	2478.4	4882.4	1712.0	3900.9	-	2051.0
	1종겸업	1270.1	4503.4	2266.0	5330.5	1532.7	8188.3	1278.5	5165.6	-	2207.6
	2종겸업	-	-	-	-	-	-	-	-	1002.2	1002.2
	전체	1207.3	3865.2	2463.5	5131.7	2363.5	5352.9	1649.0	4074.5	1002.2	1815.2
전체	전업	5991.3	5093.4	4642.5	7347.8	3349.8	5922.5	4009.2	5040.2	-	5342.2
	1종겸업	8229.2	6042.7	4712.3	7150.1	1849.5	9588.0	4235.5	6460.9	-	6876.5
	2종겸업	-	-	-	-	-	-	-	-	2827.0	2827.0
	전체	6338.2	5258.0	4655.5	7304.4	3167.5	6444.2	4097.6	5235.2	2827.0	4909.1

<표 4.5>는 영농형태별, 전·겸업별 가구당 평균 경지면적 현황이다. 대체로 1종겸업 농가의 경지면적이 전업농가의 경지면적에 비해서 큰 것으로 나타났다. 2종겸업 농가의 경지면적은 전업농가나 1종겸업 농가에 비해서 뚜렷하게 작은 것으로 나타났다. 한편, 영농형태별 구분에 따라서 살펴보면 과수, 특작, 전작 농가의 경우에는 논 면적에 비해서 상대적으로 밭 면적이 크게 나타나고 있다.

<표 4.6> 영농형태별, 전·겸업별 가구당 소득 평균(단위: 만원)

구분		조사 농가수	농가소득	농업소득	겸업소득	사업외소득	이전소득	비경상소득
논벼	전업농	737	2079.3	947.5	81.6	283.7	251.6	517.0
	1종겸업농	274	2727.6	1339.8	181.8	578.2	199.7	434.1
	전체	1011	2179.8	1008.3	97.2	329.4	243.5	504.1
과수	전업농	187	2845.6	1582.9	91.9	296.5	238.0	640.1
	1종겸업농	58	3075.3	1450.2	273.8	805.7	141.6	405.7
	전체	245	2885.5	1559.9	123.4	384.8	221.3	599.4
채소	전업농	499	2399.4	1475.0	53.9	259.2	183.5	430.8
	1종겸업농	120	3006.4	1643.1	117.5	703.6	149.4	394.9
	전체	619	2459.2	1491.6	60.2	303.0	180.1	427.2
특작	전업농	57	3252.0	2401.0	45.5	148.7	163.4	494.7
	1종겸업농	28	3558.0	2035.4	284.4	315.6	151.1	779.4
	전체	85	3319.2	2320.8	97.9	185.3	160.7	557.1
화훼	전업농	26	2175.5	1406.1	3.8	250.6	104.3	410.8
	1종겸업농	8	2011.2	556.8	-67.2	924.3	249.8	352.0
	전체	34	2155.5	1302.9	-4.9	332.4	122.0	403.6
전작	전업농	69	1792.5	847.1	99.0	269.1	183.3	394.7
	1종겸업농	20	2890.3	828.1	464.7	607.3	122.1	868.9
	전체	89	1948.8	844.4	151.1	317.2	174.6	462.2
축산	전업농	225	3402.8	2376.8	76.2	217.7	231.4	504.8
	1종겸업농	70	4079.7	2412.9	222.0	772.4	194.9	479.7
	전체	295	3499.8	2382.0	97.1	297.2	226.2	501.2
기타	전업농	15	1857.4	1025.2	78.1	333.1	138.6	311.0
	1종겸업농	5	3064.3	1457.7	39.7	1088.4	189.4	295.5
	전체	20	2023.0	1084.6	72.8	436.8	145.5	308.9
2종겸업	전체	804	3242.3	315.7	446.0	1874.0	155.6	458.9

앞선 <표 4.6>은 영농형태별, 전·겸업별 가구당 소득 평균 현황이다. 1종겸업 농가는 전업농가과 비교할 때 겸업소득, 사업외 소득에서 크게 차이가 나타나고 있음을 알 수 있다. 그러나 농가소득이나 농업소득과 비교할 때 그 차이는 상대적으로 작게 나타나고 있다. 한편, 전업농가와 1종겸업 농가는 농업소득에 대해서는 거의 차이가 나타나고 있지 않다. 다만 화훼농가인 경우에는 농업소득에서 차이가 발생하고 있고, 전작 농가의 경우에 농업소득이 다른 영농형태에 비해서 적게 나타나고 있다. 겸업소득과 사업외 소득을 고려한다면 전·겸업 구분도 반드시 필요한 변수로 판단된다.

다음 <표 4.7>은 농가소득, 농업소득, 겸업소득, 사업외 소득 변수를 각각 종속변수로 하고, 영농형태, 전·겸업, 지역(도), 경지면적, 농업종사자 수, 경영주 연령 등을 독립변수로 하는 회귀분석을 실시하여 얻은 결과이다. 전체적으로 영농형태,

전·겸업 구분, 지역, 경지면적, 농업종사자 수, 경영주의 연령 등이 소득이나 부채를 예측하는 데 중요한 설명변수인 것으로 나타났다.

<표 4.7> 2003년도 농가경제조사 결과에 대한 회귀분석 결과

(1) ANOVA for Dependent Variable 농가소득						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	
Model	20	1.108E12	5.539E10	47.05	<.0001	
Error	2963	3.488E12	1.1772E9			
Corrected Total	2983	4.596E12				
Tests of Model Effects						
Effect	Num DF	F Value	Pr > F			
Model	20	25.47	<.0001			
Intercept	1	64.64	<.0001			
sido_s (지역)	8	2.05	0.0414			
a_type (영농형태)	7	11.84	<.0001			
f_type (전겸업)	1	4.17	0.0421			
agri_num(가구원수)	1	24.51	<.0001			
age_c (가구주연령)	1	25.78	<.0001			
ag_area (경지면적)	1	50.69	<.0001			
(2) ANOVA for Dependent Variable : 농업소득						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	
Model	20	9.38E11	4.69E10	72.14	<.0001	
Error	2963	1.926E12	6.5014E8			
Corrected Total	2983	2.864E12				
Tests of Model Effects						
Effect	Num DF	F Value	Pr > F			
Model	20	29.94	<.0001			
Intercept	1	4.30	0.0390			
sido_s	8	1.38	0.2065			
a_type	7	17.17	<.0001			
f_type	1	0.02	0.8959			
agri_num	1	14.79	0.0002			
age_c	1	16.12	<.0001			
ag_area	1	68.28	<.0001			
(3) ANOVA for Dependent Variable : 농가부채						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	
Model	20	7.681E12	3.841E11	68.90	<.0001	
Error	2963	1.652E13	5.574E9			
Corrected Total	2983	2.42E13				
Tests of Model Effects						
Effect	Num DF	F Value	Pr > F			
Model	20	23.18	<.0001			
Intercept	1	5.05	0.0254			
sido_s	8	3.33	0.0012			
a_type	7	9.97	<.0001			
f_type	1	0.12	0.7306			
agri_num	1	3.09	0.0801			
age_c	1	124.52	<.0001			
ag_area	1	41.13	<.0001			(계속)

(4) ANOVA for Dependent Variable 겸업소득

Source	Sum of DF	Mean Squares	Square	F Value	Pr > F
Model	20	359.086	17.95429	16.07	<.0001
Error	2963	3309.711	1.11701		
Corrected Total	2983	3668.797			

Tests of Model Effects				
Effect	Num	DF	F Value	Pr > F
Model	20		9.70	<.0001
Intercept	1		16.27	<.0001
sido_code	8		1.81	0.0759
type	7		5.11	<.0001
kind	1		11.88	0.0007
agri_num	1		0.34	0.5601
age_c	1		19.60	<.0001
ag_area	1		3.47	0.0638

(5) ANOVA for Dependent Variable : 사업외소득

Source	Sum of DF	Mean Squares	Square	F Value	Pr > F
Model	20	5609.36	280.4678	118.02	<.0001
Error	2963	7041.51	2.3765		
Corrected Total	2983	12650.86			

Tests of Model Effects				
Effect	Num	DF	F Value	Pr > F
Model	20		33.74	<.0001
Intercept	1		53.07	<.0001
sido_code	8		4.12	0.0001
type	7		4.31	0.0002
kind	1		42.31	<.0001
agri_num	1		12.04	0.0006
age_c	1		34.63	<.0001
ag_area	1		0.43	0.5142

<표 4.8> 월별 조사에 대한 농가소득과 농업소득에 유의한 영향을 미치는 설명변수

구분	농가소득 기준	농업수입 기준
1월 조사	지역, 농업종사자수, 경영주 연령	지역, 영농형태, 전·겸업, 농업종사자수
2월 조사	농업종사자수	농업종사자수, 경지면적
3월 조사	-	영농형태
4월 조사	영농형태, 종사자수, 연령, 경지면적	지역, 영농형태, 전·겸업, 연령, 경지면적
5월 조사	농업종사자수	지역, 영농형태
6월 조사	영농형태, 전·겸업, 종사자수, 연령	지역, 영농형태, 종사자수, 연령
7월 조사	연령	연령
8월 조사	지역, 영농형태, 종사자수, 연령	지역, 영농형태, 연령
9월 조사	지역, 영농형태, 경지면적	지역, 영농형태, 경지면적
10월 조사	지역, 영농형태, 종사자수, 연령, 경지면적	지역, 영농형태, 종사자수, 경지면적
11월 조사	영농형태, 종사자수, 연령, 경지면적	영농형태, 종사자수, 경지면적
12월 조사	지역, 영농형태, 경지면적	지역, 경지면적
연단위 데이터	지역, 영농형태, 전·겸업, 종사자수, 연령, 경지면적	영농형태, 종사자수, 연령, 경지면적

주) : 유의수준 10% 이하에서 통계적으로 유의한 것을 표시함.
 : 유의수준 5% 이하에서 통계적으로 유의한 것을 표시함.
 : 유의수준 1% 이하에서 통계적으로 유의한 것을 표시함.

<표 4.8>은 월별 조사데이터에 대해서 농가소득과 농업소득에 유의한 영향을 미치는 설명변수들을 정리한 것이다. 월별로 약간의 변동은 있지만 연단위의 데이터를 분석한 결과와 비슷하게 영농형태, 전·겸업 구분, 지역, 경지면적, 농업종사자수, 경영주 연령 변수들이 소득을 예측하는 데 중요한 설명변수인 것으로 나타났다.

무응답 대체군을 만드는 과정은 증화하는 과정과 유사해서 대체군을 세분하면 대체군 내의 변수들이 동질적으로 되지만 현실적으로 적용할 때는 문제가 발생한다. 무응답 대체군을 만드는 과정은 가급적 무응답 대체군 내에서는 유사하도록 구성되 현실적인 적용측면을 고려해서 각 무응답 대체군의 크기는 어느 정도 이상이어야 한다. 무응답 대체군을 형성하는 과정은 이론과 현실의 타협이라고 볼 수 있다.

농가경제조사의 경우에도 소득 변수에 영향을 주고 있는 변수는 영농형태, 전·겸업 구분, 지역, 경지면적, 농업종사자 수, 경영주 연령 변수들로 이들 변수들을 모두 무응답 대체군을 만들 때 고려할 수는 없다. 이들 변수를 모두 고려해서는 지나치게 무응답 대체군이 세분화되어 현실적으로 적용할 수 없다. 본 연구에서는 다음의 네 가지 방안을 고려하여 이를 비교해서 최종적인 무응답 대체군 형성 방안을 제시하고자 한다.

① 방안 1

방안 1은 영농형태, 전·겸업, 지역 구분을 고려해서 일차적으로 대체군을 구성하고, 이들 대체군 중에서 표본이 큰 경우에 대해서 경지면적 변수에 대해서 의사결정나무모형을 적용하여 세분하는 방법이다.

② 방안 2

방안 2는 방안 1과 유사하게 영농형태, 전·겸업, 지역 구분을 고려해서 일차적으로 대체군을 구성하되, 지역별 변동이 적은 논벼, 2종겸업 등의 영농형태에 대해서는 지역 구분을 하지 않고, 경지면적 변수에 대해서 의사결정나무모형을 적용하여 대체군을 세분하는 방법이다.

③ 방안 3

방안 3은 방안 2와 유사하게 영농형태, 전·겸업, 지역 구분을 고려해서 일차적으로 대체군을 구성하되, 지역별 변동이 적은 논벼, 2종겸업 등의 영농형태에 대해서는 지역 구분을 하지 않고, 경지면적 변수에 대해서 의사결정분류모형을 적용하여 대체군을 세분하여 구성한다. 한편, 특작, 전작, 기타 영농형태에 대해서는 이들을 병합하고, 경지면적 변수에 대해서 의사결정나무모형을 적용하여 대체군을 세분하여 구성하는 방법이다. 방안 3은 특작, 전작, 기타는 주로 논보다는 밭을 중심의 영농형태이고, 특작, 전작, 기타 등의 영농형태는 매년 변동될 수 있다는 점을 고려할 때 이들 영농형태를 병합해서 구성할 수 있을 것이다.

④ 방안 4

방안 4는 방안 3을 기초로 하되, 논벼(전업) 농가 중에 경지면적 12,000평 미만의 표본농가로 구성된 무응답 대체군과 2종 겸업농가를 위한 무응답 대체군 중에서 대체군 내의 표본농가 수가 200 농가 이상인 대규모 대체군을 유사한 특성을 지닌 지역들로 묶어서 세분하는 방안이다.

(1) 무응답 대체군 형성 : 방안 1

무응답 대체군 형성을 위한 첫 번째 방안은 영농형태, 전·겸업, 지역 구분을 우선적으로 고려하는 방안이다. 무응답 대체군을 형성할 때 지역적 특성을 반영하기 위한 방안이다. 각 지역 구분 내에서 경지면적이나 경영주 연령 등의 변수를 이용해서 세분하여 대체군을 형성할 수 있지만 이렇게 되면 대체군의 수가 늘어나서 현실적으로 적용과정에서 어려움이 커진다.

본 연구에서는 앞선 분석 결과에 기초해서 농가경제조사의 무응답 대체군 구성을 위한 기본원칙을 다음과 같이 정하였다.

① 무응답 대체군 구성을 위해서 우선적으로 표본농가의 영농형태를 고려한다. 전국적으로 영농형태별 표본농가 수가 적은 '화훼'와 '기타' 영농형태는 각각 독립적인 무응답 대체군으로 결정한다.

② 영농형태 내의 전·겸업별 표본 농가수가 50가구 미만이 경우에는 전 영농형태 구분에 따라서 대체군을 결정한다. 이에 따라서 과수(1종겸업), 특용작물(전업), 전작(1종겸업), 전작(전업), 전작(1종겸업), 축산(1종겸업) 등은 독립적인 무응답 대체군으로 결정된다.

③ 각 도별로 충분한 표본농가가 있는 논벼(전업), 채소(전업), 2종 겸업농 등에 대해서는 각 도별로 이들 농가들을 각각 묶어서 무응답 대체군으로 구성한다. 다만, 제주도의 논벼(전업)는 경남의 논벼(전업) 대체군과 병합하여 구성한다.

④ 논벼(1종겸업), 과수(전업), 채소(1종겸업), 축산(전업)에 대해서는 표본농가 수가 적어서 지역별로 대체군을 형성할 수 없다. 본 연구에서는 이들 구분 내에서 지역, 경지면적 등을 설명변수로 하고, 농가소득을 목표변수로 하여 SAS 의사결정분류모형(decision tree model)을 적용하여 대체군을 구성하였다.

- 논벼(1종겸업)
 - (경기, 강원, 충북, 충남, 전북, 경남, 제주) & 경지면적 5,000평 미만
 - (경기, 강원, 충북, 충남, 전북, 경남, 제주) & 경지면적 5,000평 이상
 - (전남, 경북)

- 과수(전업)
 - 경기, 강원, 충북, 충남, 전북, 전남, 경남
 - 경북
 - 제주

- 채소(1종겸업)
 - 경기, 강원, 충남, 경남, 제주
 - 충북, 전북, 전남, 경북

- 축산(전업)
 - 경기, 강원, 충북, 충남, 전북, 제주
 - 전남, 경북, 경남

농가경제조사에서 사용할 무응답 대체군은 모두 44개이다. 본 연구에서 설정된 무응답 대체군 구분 현황은 다음 <표 4.9>와 같다. 부분적으로 대체군 내의 표본크기가 30개 미만인 경우에는 불가피한 경우를 제외하고 다른 유사한 대체군과 통합하여 운영할 수 있다. <표 4.10>은 2003년도 농가경제조사의 무응답 대체군별 소득변수의 평균과 표준편차를 정리한 것이다.

<표 4.9> 농가경제조사를 위한 무응답 대체군 구분 : 방안 1

대체군	영농형태	전·겸업	지역 구분
1	논벼	전업	경기
2	논벼	전업	강원
3	논벼	전업	충북
4	논벼	전업	충남
5	논벼	전업	전북
6	논벼	전업	전남
7	논벼	전업	경북
8	논벼	전업	(경남, 제주)
9	논벼	1종겸업	(경기, 강원, 충북, 충남, 전북, 경남, 제주) & 경지 5,000평 미만
10	논벼	1종겸업	(경기, 강원, 충북, 충남, 전북, 경남, 제주) & 경지 5,000평 이상
11	논벼	1종겸업	(전남, 경북)
12	과수	전업	(경기, 강원, 충북, 충남, 전북, 전남, 경남)
13	과수	전업	경북
14	과수	전업	제주
15	과수	1종겸업	전국
16	채소	전업	경기
17	채소	전업	강원
18	채소	전업	충북
19	채소	전업	충남
20	채소	전업	전북
21	채소	전업	전남
22	채소	전업	경북
23	채소	전업	경남
24	채소	전업	제주
25	채소	1종겸업	(경기, 강원, 충남, 경남, 제주)
26	채소	1종겸업	(충북, 전북, 전남, 경북)
27	특작	전업	전국
28	특작	1종겸업	전국
29	화훼	(전업, 1종겸업)	전국
30	진작	전업	전국
31	진작	1종겸업	전국
32	축산	전업	(경기, 강원, 충북, 충남, 전북, 제주)
33	축산	전업	(전남, 경북, 경남)
34	축산	1종겸업	전국
35	기타	(전업, 1종겸업)	전국
36	2종겸업		경기
37	2종겸업		강원
38	2종겸업		충북
39	2종겸업		충남
40	2종겸업		전북
41	2종겸업		전남
42	2종겸업		경북
43	2종겸업		경남
44	2종겸업		제주

<표 4.10> 2003년 농가경제조사의 무응답 대체군별 평균 및 표준편차(단위:만원)

대체군	표본크기	농가소득		농업소득	
		평균	표준편차	평균	표준편차
1	57	2381.3	1800.4	829.5	929.1
2	68	2292.3	2066.0	1017.6	1568.4
3	62	2034.4	1767.2	1007.5	1115.8
4	94	2157.0	1675.0	959.7	1128.3
5	131	2125.3	1377.5	1066.8	1136.7
6	108	2067.4	1433.6	898.0	1090.6
7	92	1810.8	1136.1	950.5	909.9
8	81	1559.0	1199.3	428.4	708.0
9	78	1674.4	880.4	672.6	387.9
10	111	3655.2	2244.2	1923.3	1566.8
11	75	2504.2	1599.4	1278.3	1381.6
12	89	2994.1	2153.0	1661.5	1836.4
13	55	2261.9	1488.4	1179.5	1060.3
14	31	2082.0	1640.1	1030.1	1365.1
15	54	3219.4	2026.8	1526.4	1474.5
16	29	2212.2	1571.8	967.2	1607.7
17	48	2299.2	2186.0	1410.8	1983.9
18	61	1826.7	1185.9	816.0	944.5
19	57	2454.2	1733.8	1325.8	1374.8
20	21	2203.3	1401.6	1268.4	1388.5
21	62	2422.1	1640.4	1611.8	1639.4
22	91	2165.4	1593.6	1332.8	1470.4
23	81	2624.8	1942.4	1727.6	1838.0
24	25	2470.3	2077.3	1493.5	1772.8
25	58	3419.7	2036.9	1887.7	1672.0
26	55	2239.1	1148.6	1086.4	928.7
27	56	2875.3	2165.1	2004.8	1977.9
28	25	3335.7	1612.6	1835.0	1686.2
29	29	2326.0	2729.4	1428.8	2927.2
30	65	1792.1	1529.9	875.3	1301.4
31	20	2285.7	2261.3	711.1	1667.6
32	111	3810.7	2637.3	2728.8	2432.0
33	90	2704.1	2180.1	1694.9	1844.9
34	61	4069.3	2446.7	2338.4	2069.7
35	19	2156.5	1222.9	1064.6	1024.3
36	131	3601.9	1896.7	363.5	590.1
37	87	2945.4	1594.1	322.1	467.7
38	70	3081.0	1511.2	273.2	459.7
39	96	3119.3	2170.8	155.3	1121.8
40	73	3096.4	1891.3	436.3	697.0
41	65	2679.9	1847.5	276.5	513.5
42	56	2863.1	1653.3	279.9	923.6
43	91	3144.9	1681.5	232.2	559.4
44	66	3843.2	1972.4	317.7	862.9

(2) 무응답 대체군 형성 : 방안 2

방안 2는 방안 1을 기초하되 각 지역별 변동이 적은 논벼(전업), 논벼(1종겸업), 2종겸업 등의 영농형태에 대해서는 지역 구분을 두지 않고, 경지면적을 고려해서 대체군을 형성하는 방안이다. 나머지 영농형태에 대해서는 방안 1과 같게 구성한다.

① 무응답 대체군 구성을 위해서 우선적으로 표본농가의 영농형태를 고려한다. 전국적으로 영농형태별 표본농가 수가 적은 '화훼'와 '기타' 영농형태는 각각 독립적인 무응답 대체군으로 결정한다.

② 영농형태 내의 전·겸업별 표본 농가수가 50가구 미만이 경우에는 영농형태 구분에 따라서 대체군을 결정한다. 이에 따라서 과수(1종겸업), 특용작물(전업), 전작(1종겸업), 전작(전업), 전작(1종겸업), 축산(1종겸업) 등은 독립적인 무응답 대체군으로 결정된다.

③ 각 지역별 변동이 적을 것으로 판단되는 논벼(전업), 논벼(1종겸업), 2종겸업 등의 영농형태에 대해서는 지역 구분을 두지 않고, 경지면적을 고려해서 의사결정 분류모형에 따라서 대체군을 형성한다.

- 논벼(전업)

- 경지면적 5,000평 미만
- 경지면적 5,000평 이상 - 12,000평 미만
- 경지면적 12,000평 이상

- 논벼(1종겸업)

- 경지면적 5,000평 미만
- 경지면적 5,000평 이상 - 12,000평 미만
- 경지면적 12,000평 이상

- 2종겸업

- (경기, 충남, 제주) & 경지면적 3,500평 미만
- (경기, 충남, 제주) & 경지면적 3,500평 이상
- (강원, 충북, 전북, 전남, 경북, 경남) & 경지면적 5,000평 미만
- (강원, 충북, 전북, 전남, 경북, 경남) & 경지면적 5,000평 이상

④ 지역별 특성을 고려해야 하는 채소(전업)에 대해서는 각 도별로 이들 농가들을 각각 묶어서 무응답 대체군으로 구성한다.

⑤ 과수(전업), 채소(1종겸업), 축산(전업)에 대해서는 표본농가 수가 적어서 지역별로 대체군을 형성할 수 없다. 이들에 대한 구분은 방안 1을 따른다. 방안 2에 의해서 구성되는 무응답 대체군은 모두 34개이다.

<표 4.11> 농가경제조사를 위한 무응답 대체군 구분 : 방안 2

대체군	영농형태	전겸업	지역 구분
1	논벼	전업	경지면적 5,000평 미만
2	논벼	전업	경지면적 5,000평 이상 - 12,000평 미만
3	논벼	전업	경지면적 12,000평 이상
4	논벼	1종겸업	경지면적 5,000평 미만
5	논벼	1종겸업	경지면적 5,000평 이상 - 12,000평 미만
6	논벼	1종겸업	경지면적 12,000평 이상
7	과수	전업	(경기, 강원, 충북, 충남, 전북, 전남, 경남)
8	과수	전업	경북
9	과수	전업	제주
10	과수	1종겸업	전국
11	채소	전업	경기
12	채소	전업	강원
13	채소	전업	충북
14	채소	전업	충남
15	채소	전업	전북
16	채소	전업	전남
17	채소	전업	경북
18	채소	전업	경남
19	채소	전업	제주
20	채소	1종겸업	(경기, 강원, 충남, 경남, 제주)
21	채소	1종겸업	(충북, 전북, 전남, 경북)
22	특작	전업	전국
23	특작	1종겸업	전국
24	화훼	(전업, 1종겸업)	전국
25	진작	전업	전국
26	진작	1종겸업	전국
27	축산	전업	(경기, 강원, 충북, 충남, 전북, 제주)
28	축산	전업	(전남, 경북, 경남)
29	축산	1종겸업	전국
30	기타	(전업, 1종겸업)	전국
31	2종겸업		(경기, 충남, 제주) & 경지면적 3,500평 미만
32	2종겸업		(경기, 충남, 제주) & 경지면적 3,500평 이상
33	2종겸업		(강원, 충북, 전북, 전남, 경북, 경남) & 경지 5,000평 미만
34	2종겸업		(강원, 충북, 전북, 전남, 경북, 경남) & 경지 5,000평 이상

<표 4.12> 농가경제조사의 무응답 대체군별 평균 및 표준편차(단위:만원) : 방안2

대체군	표본크기	농가소득		농업소득	
		평균	표준편차	평균	표준편차
1	407	1561.53	1168.13	440.98	463.53
2	235	2373.81	1403.37	1287.65	1134.81
3	51	4346.59	2233.09	2861.89	1651.53
4	108	1684.95	877.57	636.99	412.14
5	104	2710.41	1168.54	1322.63	730.17
6	52	5005.45	2724.15	2989.89	2112.65
7	89	2994.10	2152.96	1661.45	1836.38
8	55	2261.85	1488.44	1179.45	1060.34
9	31	2081.97	1640.11	1030.07	1365.05
10	54	3219.42	2026.79	1526.41	1474.45
11	29	2212.16	1571.75	967.24	1607.71
12	48	2299.18	2185.99	1410.78	1983.90
13	61	1826.70	1185.90	815.97	944.53
14	57	2454.17	1733.83	1325.78	1374.82
15	21	2203.33	1401.58	1268.39	1388.53
16	62	2422.08	1640.36	1611.79	1639.44
17	91	2165.35	1593.63	1332.75	1470.44
18	81	2624.84	1942.44	1727.60	1838.04
19	25	2470.28	2077.27	1493.53	1772.76
20	74	3205.84	1964.47	1746.07	1550.58
21	39	2160.52	992.69	1026.49	963.94
22	56	2875.26	2165.06	2004.81	1977.89
23	25	3335.65	1612.57	1835.01	1686.24
24	29	2326.01	2729.42	1428.76	2927.22
25	65	1792.05	1529.93	875.33	1301.38
26	20	2285.72	2261.32	711.13	1667.57
27	111	3810.69	2637.33	2728.80	2431.96
28	90	2704.12	2180.14	1694.85	1844.87
29	61	4069.31	2446.67	2338.39	2069.74
30	19	2156.48	1222.87	1064.57	1024.28
31	194	3268.32	1672.54	242.07	554.82
32	99	3948.42	2517.36	368.96	1258.54
33	396	2907.27	1565.25	257.34	407.38
34	46	3639.00	2485.68	692.87	1405.44

(3) 무응답 대체군 형성 : 방안 3

방안 3은 방안 2를 기초로 하되, 전작, 기타의 영농형태를 합병해서 처리하는 방안이다. 전작, 기타 영농형태의 표본농가 수는 전체적으로 작고, 전작, 기타는 주로 논보다는 밭을 중심의 영농형태이며, 이러한 영농형태는 매년 변동될 수 있다는 점을 고려할 때 이들 영농형태를 병합해서 구성할 수 있을 것이다. 앞선 <표 4.6>의 분석결과에서도 전작과 기타 농가는 비슷한 소득 형태를 갖는 것으로 나타났다.

즉, 방안 3은 전작, 기타 영농형태에 대한 대체군 형성방법이 달리질 뿐 나머지 영농형태에 대한 대체군은 방안 2와 동일하다. 방안 3에 의해서 구성되는 무응답 대체군은 모두 33개이다.

- 전작, 기타(전업)
 - 전국

- 전작, 기타(1종겸업)
 - 전국

<표 4.13> 농가경제조사를 위한 무응답 대체군 구분 : 방안 3

대체군	영농형태	전겸업	지역 구분
1	논벼	전업	경지면적 5,000평 미만
2	논벼	전업	경지면적 5,000평 이상 - 12,000평 미만
3	논벼	전업	경지면적 12,000평 이상
4	논벼	1종겸업	경지면적 5,000평 미만
5	논벼	1종겸업	경지면적 5,000평 이상 - 12,000평 미만
6	논벼	1종겸업	경지면적 12,000평 이상
7	과수	전업	(경기, 강원, 충북, 충남, 전북, 전남, 경남)
8	과수	전업	경북
9	과수	전업	제주
10	과수	1종겸업	전국
11	채소	전업	경기
12	채소	전업	강원
13	채소	전업	충북
14	채소	전업	충남
15	채소	전업	전북
16	채소	전업	전남
17	채소	전업	경북
18	채소	전업	경남
19	채소	전업	제주
20	채소	1종겸업	(경기, 강원, 충남, 경남, 제주)
21	채소	1종겸업	(충북, 전북, 전남, 경북)
22	특작	전업	전국
23	특작	1종겸업	전국
24	화훼	(전업, 1종겸업)	전국
25	전작, 기타	전업	전국
26	전작, 기타	1종겸업	전국
27	축산	전업	(경기, 강원, 충북, 충남, 전북, 제주)
28	축산	전업	(전남, 경북, 경남)
29	축산	1종겸업	전국
30	2종겸업		(경기, 충남, 제주) & 경지면적 3,500평 미만
31	2종겸업		(경기, 충남, 제주) & 경지면적 3,500평 이상
32	2종겸업		(강원, 충북, 전북, 전남, 경북, 경남) & 경지 5,000평 미만
33	2종겸업		(강원, 충북, 전북, 전남, 경북, 경남) & 경지 5,000평 이상

<표 4.14> 농가경제조사의 무응답 대체군별 평균 및 표준편차(단위:만원) : 방안 3

대체군	표본크기	농가소득		농업소득	
		평균	표준편차	평균	표준편차
1	407	1561.53	1168.13	440.98	463.53
2	235	2373.81	1403.37	1287.65	1134.81
3	51	4346.59	2233.09	2861.89	1651.53
4	108	1684.95	877.57	636.99	412.14
5	104	2710.41	1168.54	1322.63	730.17
6	52	5005.45	2724.15	2989.89	2112.65
7	89	2994.10	2152.96	1661.45	1836.38
8	55	2261.85	1488.44	1179.45	1060.34
9	31	2081.97	1640.11	1030.07	1365.05
10	54	3219.42	2026.79	1526.41	1474.45
11	29	2212.16	1571.75	967.24	1607.71
12	48	2299.18	2185.99	1410.78	1983.90
13	61	1826.70	1185.90	815.97	944.53
14	57	2454.17	1733.83	1325.78	1374.82
15	21	2203.33	1401.58	1268.39	1388.53
16	62	2422.08	1640.36	1611.79	1639.44
17	91	2165.35	1593.63	1332.75	1470.44
18	81	2624.84	1942.44	1727.60	1838.04
19	25	2470.28	2077.27	1493.53	1772.76
20	74	3205.84	1964.47	1746.07	1550.58
21	39	2160.52	992.69	1026.49	963.94
22	56	2875.26	2165.06	2004.81	1977.89
23	25	3335.65	1612.57	1835.01	1686.24
24	29	2326.01	2729.42	1428.76	2927.22
25	79	1795.76	1433.84	879.67	1233.65
26	25	2452.21	2141.06	874.68	1604.31
27	111	3810.69	2637.33	2728.80	2431.96
28	90	2704.12	2180.14	1694.85	1844.87
29	61	4069.31	2446.67	2338.39	2069.74
30	194	3268.32	1672.54	242.07	554.82
31	99	3948.42	2517.36	368.96	1258.54
32	396	2907.27	1565.25	257.34	407.38
33	46	3639.00	2485.68	692.87	1405.44

(4) 무응답 대체군 형성 : 방안 4

방안 4는 방안 3을 기초로 하여 전작, 기타의 영농형태를 합병해서 처리하고, 논벼(전업) 농가 중에 경지면적 12,000평 미만의 표본농가로 구성된 무응답 대체군과 2종 겸업농가를 위한 무응답 대체군 중에서 대체군 내의 표본농가의 수가 200 농가 이상인 대체군을 유사한 성격의 지역으로 세분하는 방안이다.

- 논벼, 전업, 경지면적 5,000평 미만
 - 경기, 강원, 충남
 - 충북, 경북, 제주
 - 전북, 전남, 경남

- 논벼, 전업, 경지면적 5,000평 이상 - 12,000평 미만
 - 경기, 강원, 충남
 - 충북, 전북, 경북, 제주
 - 전남, 경남

- 2종겸업, (강원, 충북, 전북, 전남, 경북, 경남) & 경지 5,000평 미만
 - 강원, 충북
 - 전북, 전남
 - 경북, 경남

위의 과정을 거쳐서 최종적으로 방안 4에 의해서 구성되는 무응답 대체군은 모두 39개이다.

<표 4.15> 농가경제조사를 위한 무응답 대체군 구분 : 방안 4

대체군	영농형태	전겸업	지역 구분
1	논벼	전업	경지 5,000평 미만 & (경기, 강원, 충남)
2	논벼	전업	경지 5,000평 미만 & (충북, 경북, 제주)
3	논벼	전업	경지 5,000평 미만 & (전북, 전남, 경남)
4	논벼	전업	경지 5,000평 이상 - 12,000평 미만 & (경기, 강원, 충남)
5	논벼	전업	경지 5,000평 이상 - 12,000평 미만 & (충북, 전북, 경북, 제주)
6	논벼	전업	경지 5,000평 이상 - 12,000평 미만 & (전남, 경남)
7	논벼	전업	경지 12,000평 이상
8	논벼	1종겸업	경지 5,000평 미만
9	논벼	1종겸업	경지 5,000평 이상 - 12,000평 미만
10	논벼	1종겸업	경지 12,000평 이상
11	과수	전업	(경기, 강원, 충북, 충남, 전북, 전남, 경남)
12	과수	전업	경북
13	과수	전업	제주
14	과수	1종겸업	전국
15	채소	전업	경기
16	채소	전업	강원
17	채소	전업	충북
18	채소	전업	충남
19	채소	전업	전북
20	채소	전업	전남
21	채소	전업	경북
22	채소	전업	경남
23	채소	전업	제주
24	채소	1종겸업	(경기, 강원, 충남, 경남, 제주)
25	채소	1종겸업	(충북, 전북, 전남, 경북)
26	특작	전업	전국
27	특작	1종겸업	전국
28	화훼	(전업, 1종겸업)	전국
29	전작, 기타	전업	전국
30	전작, 기타	1종겸업	전국
31	축산	전업	(경기, 강원, 충북, 충남, 전북, 제주)
32	축산	전업	(전남, 경북, 경남)
33	축산	1종겸업	전국
34	2종겸업		(경기, 충남, 제주) & 경지 3,500평 미만
35	2종겸업		(경기, 충남, 제주) & 경지 3,500평 이상
36	2종겸업		(강원, 충북) & 경지 5,000평 미만
37	2종겸업		(전북, 전남) & 경지 5,000평 미만
38	2종겸업		(경북, 경남) & 경지 5,000평 미만
39	2종겸업		(강원, 충북, 전북, 전남, 경북, 경남) & 경지 5,000평 이상

<표 4.16> 농가경제조사의 무응답 대체군별 평균 및 표준편차(단위:만원) : 방안 4

대체군	표본크기	농가소득		농업소득	
		평균	표준편차	평균	표준편차
1	127	1769.75	1394.09	467.40	394.69
2	79	1279.75	978.72	364.76	623.01
3	201	1540.72	1054.83	454.25	428.64
4	74	2476.54	1599.28	1248.82	1332.82
5	109	2364.72	1426.62	1452.57	1039.47
6	52	2246.70	1016.55	997.21	965.18
7	51	4346.59	2233.09	2861.89	1651.53
8	108	1684.95	877.57	636.99	412.14
9	104	2710.41	1168.54	1322.63	730.17
10	52	5005.45	2724.15	2989.89	2112.65
11	89	2994.10	2152.96	1661.45	1836.38
12	55	2261.85	1488.44	1179.45	1060.34
13	31	2081.97	1640.11	1030.07	1365.05
14	54	3219.42	2026.79	1526.41	1474.45
15	29	2212.16	1571.75	967.24	1607.71
16	48	2299.18	2185.99	1410.78	1983.90
17	61	1826.70	1185.90	815.97	944.53
18	57	2454.17	1733.83	1325.78	1374.82
19	21	2203.33	1401.58	1268.39	1388.53
20	62	2422.08	1640.36	1611.79	1639.44
21	91	2165.35	1593.63	1332.75	1470.44
22	81	2624.84	1942.44	1727.60	1838.04
23	25	2470.28	2077.27	1493.53	1772.76
24	74	3205.84	1964.47	1746.07	1550.58
25	39	2160.52	992.69	1026.49	963.94
26	55	1426.20	1190.49	582.60	1155.52
27	27	3726.35	2566.34	2551.34	2212.32
28	53	2336.35	1467.76	1525.18	1413.31
29	50	2893.95	1928.22	1354.54	1699.67
30	29	2326.01	2729.42	1428.76	2927.22
31	111	3810.69	2637.33	2728.80	2431.96
32	90	2704.12	2180.14	1694.85	1844.87
33	61	4069.31	2446.67	2338.39	2069.74
34	194	3268.32	1672.54	242.07	554.82
35	99	3948.42	2517.36	368.96	1258.54
36	137	2994.51	1474.54	277.70	353.20
37	125	2773.50	1731.09	292.70	480.11
38	134	2942.86	1494.39	203.54	381.97
39	46	3639.00	2485.68	692.87	1405.44

(5) 무응답 대체군 형성 방안의 비교

무응답 대체의 효율성을 우선적으로 고려한다면 가급적 대체군을 세분하는 것이 바람직하지만, 농가의 영농형태는 매년 변동하기 때문에 무응답 대체군을 세분하게 되면 어느 해에는 대체군에 속한 표본수가 대단히 적거나 없는 경우가 발생할 수 있다. 따라서 무응답 대체군을 형성할 때는 무응답 대체의 효율성 측면과 안정성 측면을 함께 고려해야 한다.

앞서 제시한 네 가지의 무응답 대체군 작성 방안을 비교하자. 먼저, 무응답 대체군을 층으로 간주하고, 무응답의 발생 빈도는 대체군의 크기에 비례한다고 가정하자. 이와 같은 경우에 대체군을 형성하여 무응답을 대체하는 경우와 대체군을 형성하지 않고 전체 조사 데이터에 대하여 핫덱방법을 적용하는 경우를 비교해 보자.

대체군 형성과정이 효과적이라면 대체군 내의 조사값들이 동질적이어서 대체군 내 분산()이 작을 것이다. 각 변수에 대한 대체군 형성 효과는 대체군을 층으로 간주하여 비례배정을 하는 경우에 설계효과를 계산하여 살펴볼 수 있다. 무응답 대체군 형성 효과는 다음과 같이 정의할 수 있다.

① 전체를 하나의 대체군으로 간주하는 경우의 무응답 대체에 따른 변동 크기

여기서 n_0 : 무응답 농가수, N : 전체 표본 농가수

② 무응답 대체군을 구성하는 경우의 무응답 대체에 따른 변동 크기

여기서 n_0 : 무응답 농가수, N : 전체 표본수, n_k : 각 대체군에서 표본크기

③ 무응답 대체군 형성의 효과

무응답 대체군 형성 효과

각 방안별 대체군 형성 효과는 중요한 변수에 대해서 앞서 정의한 ‘무응답 대체군 효과’를 계산하여 살펴 볼 수 있다. 무응답 대체군 형성 효과는 백분율로 나타내어 그 값이 커질수록 무응답 대체군이 효과적으로 구성되었다는 것을 의미한다. 본 연구에서는 네 가지 방안을 비교하기 위하여 고려하는 주요변수는 농가소득, 농업소득, 농가부채, 겸업소득, 사업외 소득 등이다.

<표 4.17> 각 방안별 주요변수에 대한 무응답 대체군 형성 효과 비교

구분	농가소득	농업소득	겸업소득	사업외 소득	농가부채
방안1	10.5%	17.5%	8.7%	37.9%	6.1%
방안2	15.8%	23.8%	9.4%	36.9%	12.4%
방안3	15.9%	24.0%	9.5%	36.9%	12.4%
방안4	16.3%	24.2%	9.4%	37.1%	12.5%

연간 데이터를 이용해서 분석할 때 방안 4가 가장 효과적인 것으로 나타났다. 방안 4를 이용하여 대체군을 형성하게 되면 무응답 대체로 인한 변동을 농가소득 변수에 대해서는 약 16%, 농업소득 변수는 약 24%, 겸업소득 변수는 10%, 사업외 소득 변수는 약 37%, 농가부채 변수는 약 13% 정도 낮출 수 있다.

이러한 분석 결과를 종합해 볼 때 농가경제조사를 위한 무응답 대체군 형성방법은 경지면적 변수를 대체군 형성에 사용하는 방안 2, 방안 3, 방안 4 등을 사용할 수 있다. 본 연구에서는 최종적인 무응답 대체군 형성방안으로 방안 4를 제안한다.

4.3 어가경제조사를 위한 대체군

어가경제조사의 대체군 설정과 관련하여 고려한 주요 관심변수는 어가소득, 어업소득 변수이고, 이들 변수에 영향을 줄 것으로 예상되는 설명변수로는 어업형태, 전·겸업, 지역, 선박톤수, 양식장 면적 등을 고려하였다.

다음 <표 4.18>은 2003년도 어가경제조사에 대해서 전국의 어업형태 및 전·겸업별 표본크기 현황이다. 어업형태와 전·겸업 구분에 따라서 대단한 편차가 있음을

알 수 있다. 무동력선 사용어가는 3가구로 나타났다.

한편, <표 4.19>은 각 지역별 어업형태 및 전·겸업 구분별 표본크기 현황을 정리한 것이다. 지역적으로 동력어선 사용어가는 비교적 고르게 분포하고 있지만 나머지 어업형태는 지역적으로 편중되어 있다.

<표 4.18> 전국의 어업형태 및 전·겸업별 표본크기 현황

	비어선 어가	무동력선 어가	동력선 어가	양식업	전체
전업어가	8	1	182	82	273
1종겸업 어가	23	1	329	125	478
2종겸업 어가	124	1	201	98	424
전체	155	3	712	305	1175

<표 4.19> 지역 및 어업형태별 표본크기 현황

구 분	전·겸업	비어선 사용 어가	무동력선 사용 어가	동력선 사용 어가	양식업 어 가	전체
전남, 전북	전업어가	1	1	55	46	103
	1종겸업 어가	5	0	79	67	151
	2종겸업 어가	18	0	45	35	98
	전체	24	1	179	148	352
경남, 경북, 충남	전업어가	4	0	70	22	96
	1종겸업 어가	5	1	133	40	179
	2종겸업 어가	16	0	76	29	121
	전체	25	1	279	91	396
부산, 인천, 경기	전업어가	0	0	14	7	21
	1종겸업 어가	8	0	25	6	39
	2종겸업 어가	23	0	22	7	52
	전체	31	0	61	20	112
울산, 강원, 제주	전업어가	3	0	40	0	43
	1종겸업 어가	4	0	77	3	84
	2종겸업 어가	20	0	35	2	57
	전체	27	0	152	5	184
전업, 1종 규모 이하 층	전업어가	0	0	0	4	4
	1종겸업 어가	0	0	9	4	13
	2종겸업 어가	15	1	1	11	28
	전체	15	1	10	19	45
2종 겸업 층	전업어가	0	0	3	3	6
	1종겸업 어가	1	0	6	5	12
	2종겸업 어가	32	0	22	14	68
	전체	33	0	31	22	86

다음 <표 4.20>는 어업형태별, 전·겸업별 가구당 어선 및 어장 보유비율과 가구당 평균 어선톤수와 평균 양식장 면적 평균 현황이다. 전체적으로 전업 어가의 어선 보유비율이 1종겸업 어가나 전업어가의 비율보다 높게 나타났지만, 큰 차이를 보이고 있지는 않다. 가구당 평균 어선 톤수와 어장 면적도 전업 어가의 값이 1종겸업이나 2종겸업 어가에 비해서 크게 나타났다.

<표 4.20> 어업형태별, 전·겸업별 가구당 어선 및 어장 보유비율과 평균

구 분		조사 어가수	어선 사용 어가비율	어가당 평균 어선 톤수	양식장 사용 어가 비율	어가당 평균 양식장 면적
비어선 어가	전업어가	8	45.0%	1.75	36.9%	0.22
	1종겸업	23	11.0%	0.20	20.7%	0.13
	2종겸업	124	13.0%	0.25	5.4%	0.04
	계	155	13.3%	0.27	7.5%	0.05
무동력 또는 동력선 어가	전업어가	174	97.0%	4.75	9.1%	0.04
	1종겸업	318	95.0%	3.37	20.4%	0.20
	2종겸업	195	92.0%	2.69	22.0%	0.18
	계	687	94.3%	3.42	18.6%	0.16
양식업 어가	전업어가	73	87.9%	2.24	100.0%	4.09
	1종겸업	119	96.3%	2.42	100.0%	3.75
	2종겸업	93	65.3%	1.12	100.0%	1.74
	계	285	81.9%	1.86	100.0%	3.03

<표 4.21> 어업형태별, 전·겸업별 가구당 소득 평균(단위: 만원)

구 분		조사어가수	어가소득	어업소득	겸업소득	사업외소득	이전소득	비경상소득
비어선 어가	전업	8	1272.5	977.2	36.2	93.2	86.1	89.8
	1종겸업	23	1489.0	810.4	94.5	186.6	95.1	302.3
	2종겸업	124	2247.9	399.3	429.4	789.9	101.3	528.0
	계	155	2157.8	449.0	390.4	719.8	104.0	498.2
무동력선 또는 동력선 사용어가	전업	174	2681.8	2083.6	109.7	68.0	72.1	348.5
	1종겸업	318	2579.6	1631.4	202.4	271.0	81.6	393.2
	2종겸업	195	2852.4	578.4	838.9	796.4	80.8	558.6
	계	687	2698.3	1351.7	409.3	415.3	79.3	442.7
양식업 어가	전업	73	3036.8	2690.3	7.0	73.5	24.9	241.1
	1종겸업	119	2707.3	1964.8	125.7	209.5	74.6	332.7
	2종겸업	93	2379.7	511.3	867.7	586.0	130.2	284.5
	계	285	2656.1	1560.9	392.2	326.6	84.8	291.6

<표 4.21>은 어업형태별, 전·겸업별 가구당 소득 평균 현황이다. 어업소득 측면에서 비교하면 전업어거나 1종겸업 어가에 비해서 2종겸업 어가의 어업소득이 아주 작게 나타나고 있다. 한편, 겸업소득, 사업외 소득 측면에서 보면 2종겸업 어가의 소득이 전업어거나 1종겸업어가에 비해서 큰 것으로 나타났다. 어업형태별 구분에 따라서 소득의 차이를 보면 양식업 어가의 어업소득이 가장 많은 것으로 나타났고, 상대적으로 비어선 또는 무동력선 사용 어가의 소득이 적은 것으로 나타났다.

다음 <표 4.22>은 어가소득, 어업소득 변수를 각각 종속변수로 하고, 어업형태, 전·겸업, 지역(도) 등을 독립변수로 하는 회귀분석을 실시하여 얻은 결과이다. 전체적으로 어업형태, 전·겸업 구분, 지역, 선박톤수, 어장면적 등의 변수들이 소득을 예측하는 데 중요한 설명변수인 것으로 나타났다.

<표 4.22> 2003년도 어가경제조사 결과에 대한 회귀분석 결과

(1) ANOVA for Dependent Variable 어가소득					
Source	DF	Squares	Sum of Square	Mean Square	Pr > F
Model	12	2.164E10	1.8033E9	6.70	<.0001
Error	1114	2.998E11	2.6914E8		
Corrected Total	1126	3.215E11			

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
Intercept	1	357.73	<.0001
st_d (지역)	5	0.46	0.8068
type (어업형태)	3	8.52	<.0001
kind (전겸업)	2	1.05	0.3515
f_area (어장면적)	1	3.87	0.0501
s_ton (어선톤수)	1	18.71	<.0001

(2) ANOVA for Dependent Variable 어업소득					
Source	DF	Squares	Sum of Square	Mean Square	Pr > F
Model	12	5.593E10	4.6607E9	32.09	<.0001
Error	1114	1.618E11	1.4525E8		
Corrected Total	1126	2.177E11			

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
Intercept	1	44.80	<.0001
st_d	5	1.74	0.1238
type	3	1.98	0.1161
kind	2	58.56	<.0001
f_area	1	4.18	0.0418
s_ton	1	14.94	0.0001

(3) ANOVA for Dependent Variable 겸업소득

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	9.84E9	8.2E8	20.79	<.0001
Error	1114	4.393E10	39436850		
Corrected Total	1126	5.377E10			

Tests of Model Effects

Effect	Num DF	F Value	Pr > F
Intercept	1	30.47	<.0001
st_d	5	2.34	0.0413
type	3	3.11	0.0264
kind	2	43.11	<.0001
f_area	1	0.61	0.4371
s_ton	1	8.84	0.0032

(4) ANOVA for Dependent Variable 사업외소득

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	7.2856E9	6.0714E8	21.03	<.0001
Error	1114	3.217E10	28876293		
Corrected Total	1126	3.945E10			

Tests of Model Effects

Effect	Num DF	F Value	Pr > F
Intercept	1	87.48	<.0001
st_d	5	2.43	0.0347
type	3	0.98	0.4022
kind	2	59.36	<.0001
f_area	1	0.16	0.6853
s_ton	1	0.19	0.6619

다음 <표 4.23>은 월별 조사데이터에 대해서 어가소득, 어업소득 변수를 각각 종속변수로 하고, 어업형태, 전·겸업, 지역(도), 선박톤수, 어장면적 등을 독립변수로 하는 회귀분석을 실시하여 얻은 결과이다. 월별 변동은 있지만 전체적으로 어업형태, 전·겸업 구분, 지역 변수들이 소득을 예측하는 데 중요한 설명변수인 것으로 나타났다.

따라서 대체군 결정과정에서는 우선적으로 어업형태 변수를 1차적으로 고려하여 대체군을 만들고, 이후에 전·겸업 구분과 지역을 고려해서 대체군을 만드는 것이 바람직하다.

<표 4.23> 월별 조사데이터에서 소득 관련 변수에 유의한 영향을 주는 설명변수

구분	어가소득 기준	어업수입 기준
1월 조사	어선톤수	어업형태, 전·겸업, 어선톤수
2월 조사	어업형태, 어장면적	어업형태, 전·겸업, 어장면적
3월 조사	어업형태, 어선톤수	어업형태, 전·겸업
4월 조사	어선톤수, 어장면적	전·겸업, 어선톤수, 어장면적
5월 조사	어업형태, 어선톤수	전·겸업, 어선톤수
6월 조사	-	전·겸업
7월 조사	지역, 어업형태	지역, 전·겸업
8월 조사	지역, 어업형태	지역, 전·겸업
9월 조사	어업형태	어업형태
10월 조사	지역, 어선톤수, 어장면적	지역, 전·겸업, 어선톤수, 어장면적
11월 조사	어업형태, 어선톤수, 어장면적	전·겸업, 어선톤수, 어장면적
12월 조사	어선톤수	어업형태, 전·겸업, 어선톤수
연단위 데이터	어업형태, 어선톤수, 어장면적	전·겸업, 어선톤수, 어장면적

주) : 유의수준 10% 이하에서 통계적으로 유의한 것을 표시함.
 : 유의수준 5% 이하에서 통계적으로 유의한 것을 표시함.
 : 유의수준 1% 이하에서 통계적으로 유의한 것을 표시함.

어가경제조사에 대한 무응답 대체군은 농가경제조사의 무응답 대체군 형성과 유사한 방법을 사용한다. 본 연구에서는 다음의 네 가지 방안을 고려하여 이를 비교해서 최종적인 무응답 대체군 형성 방안을 제시하고자 한다.

① 방안 1

방안 1은 우선적으로 어가의 어업형태와 전·겸업 구분을 고려하여 무응답 대체군을 형성하는 방안이다. 지역은 표본설계 상의 지역층(몇 개의 도를 묶어서 구성되었음)을 고려하여 유사한 특성을 나타내는 지역끼리 묶어서 무응답 대체군을 구성하였다.

② 방안 2

방안 2는 방안 1과 마찬가지로 어가의 어업형태와 전·겸업 구분을 우선적으로 고려하여 무응답 대체군을 형성하는 방안이다. 다만, 지역은 행정구역상의 각 시·도 구분을 고려하여 유사한 특성을 나타내는 지역끼리 묶어서 무응답 대체군을 구성하였다.

③ 방안 3

방안 3은 방안 1과 마찬가지로 어가의 어업형태와 전·겸업 구분을 우선적으로 고려하여 무응답 대체군을 형성하는 방안이다. 다만, 동력선 사용어가와 양식업 어가에 대해서 선박톤수와 양식장 면적을 무응답 대체군 형성과정에서 이용하고, 이후에 지역을 대체군 세분 과정에서 이용하는 방안이다.

④ 방안 4

방안 4은 방안 3과 마찬가지로 어가의 어업형태와 전·겸업 구분을 우선적으로 고려하여 무응답 대체군을 형성하는 방안이다. 다만, 지역을 우선적으로 고려하여 대체군을 형성하고, 이후에 선박톤수와 양식장 면적을 무응답 대체군 세분하는 과정에서 이용하는 방안이다.

(1) 방안 1

방안 1에서는 <표 4.18>부터 <표 4.23>의 분석결과에 기초해서 어가경제조사를 위한 무응답 대체군 구성의 기본원칙을 다음과 같이 정하였다.

① 무응답 대체군 구성을 위해서 우선적으로 표본어가의 어업형태를 고려한다. 전국적으로 어업형태별 표본농가가 3가구인 '무동력선 사용어가'는 '동력선 사용 어가'에 병합하여 대체군을 구성한다.

② 지역은 세부적인 층 구분이 아닌 지역층(몇 개의 도)을 고려한다.

③ 전체 조사결과를 '비어선 어가', '무동력선 및 동력선 사용어가', '양식업 어가'로 구분하여 각각의 어업형태에 대해서 어가소득 변수를 목표변수로 하고, 전·겸업 구분과 지역 구분을 설명변수로 하여 SAS의 의사결정분류모형(decision tree

model)을 적용하여 대체군을 구성하였다.

다음 <표 4.24>은 본 연구에서 제시하는 어가경제조사를 위한 무응답 대체군 현황이다. 어가경제조사에서 무응답 대체를 위해서 사용되는 대체군은 전체적으로 12개이다. 한편, <표 4.25>는 설정된 무응답 대체군별 어가소득과 어업소득 변수에 대해서 평균값과 표준편차를 구하여 정리한 것이다.

<표 4.24> 어가경제조사를 위한 무응답 대체군 구분 : 방안 1

대체군	어업형태	전·겸업	지역(층번호)
1	비어선 어가	전업, 1종겸업	전체
2	비어선 어가	2종겸업	전체
3	무동력선, 동력선 사용 어가	전업어가	31, 32, 33
4	무동력선, 동력선 사용 어가	전업어가	1-5, 11-17, 21-23, 50, 90
5	무동력선, 동력선 사용 어가	1종겸업	1-5, 21-23
6	무동력선, 동력선 사용 어가	1종겸업	11-17, 50, 90
7	무동력선, 동력선 사용 어가	2종겸업	1-5, 21-23, 50
8	무동력선, 동력선 사용 어가	2종겸업	11-17, 90
9	무동력선, 동력선 사용 어가	1, 2종겸업	31, 32, 33
10	양식업	전업	전체
11	양식업	1종겸업	전체
12	양식업	2종겸업	전체

<표 4.25> 어가경제조사의 무응답 대체군별 평균 및 표준편차(단위:만원) : 방안 1

대체군	표본크기	어가소득		어업소득	
		평균	표준편차	평균	표준편차
1	31	1442.61	734.80	844.38	568.36
2	124	2155.92	1742.15	406.31	506.49
3	36	2771.91	1563.66	2299.50	1446.11
4	138	2591.84	2205.19	2105.48	2206.94
5	100	2678.23	1903.53	1678.31	1335.10
6	142	2661.33	2121.18	1801.03	1819.57
7	68	2422.02	1771.29	587.38	559.17
8	92	2843.6	2178.19	577.78	812.57
9	111	2842.43	1882.55	1517.19	1394.98
10	73	3130.11	3091.57	2789.09	3097.78
11	119	2900.83	2468.22	2133.41	2312.06
12	93	2325.38	1416.70	546.30	664.79

(2) 방안 2

방안 2는 방안 1과 마찬가지로 어가의 어업형태와 전·겸업 구분을 우선적으로 고려하여 무응답 대체군을 형성하는 방안이다. 다만, 지역은 행정구역상의 각 시·도 구분을 고려하여 유사한 특성을 지닌 지역끼리 묶어서 무응답 대체군을 구성하였다.

다음 <표 4.26>는 어가경제조사의 무응답 대체군 형성을 위해서 유사한 성격을 지니는 인근 시·도를 묶어서 구성한 지역별, 어업형태별, 전·겸업별 표본어가 수 현황이다. 방안 2에서는 표본설계상의 지역 구분 대신에 <표 4.26>에 제시된 지역 구분을 이용하여 무응답 대체군을 구성한 것이다.

<표 4.26> 각 지역별, 어업형태별, 전·겸업별 표본어가 수 현황

지역	비어선 어가			무동력선, 동력선 어가			양식업 어가			합계
	전업	1종겸업	2종겸업	전업	1종겸업	2종겸업	전업	1종겸업	2종겸업	
경기, 인천, 충남	1	5	30	7	37	28	4	10	31	153
강원, 경북	4	6	10	49	88	45	0	0	3	205
전북, 전남, 제주	3	8	66	66	102	61	46	71	40	464
부산, 울산, 경남	0	4	18	52	91	60	23	38	19	305
전국	9	23	125	173	318	194	73	119	93	1127

다음 <표 4.27>은 방안 2의 무응답 대체군 현황이다. 어가경제조사에서 무응답 대체를 위해서 사용되는 대체군은 전체적으로 20개이다. 한편, <표 4.28>은 설정된 무응답 대체군별 어가소득과 어업소득 변수에 대해서 평균값과 표준편차를 구하여 정리한 것이다.

<표 4.27> 어가경제조사를 위한 무응답 대체군 구분 : 방안 2

대체군	어업형태	전겸업	지역 구분
1	비어선 어가	전업, 1종겸업	전체
2	비어선 어가	2종겸업	전북, 전남, 제주
3	비어선 어가	2종겸업	전북, 전남, 제주 제외 나머지
4	무동력선, 동력선 사용 어가	전업어가	경기, 인천, 충남, 강원, 경북
5	무동력선, 동력선 사용 어가	전업어가	전북, 전남, 제주
6	무동력선, 동력선 사용 어가	전업어가	부산, 울산, 경남
7	무동력선, 동력선 사용 어가	1종겸업	경기, 인천, 충남
8	무동력선, 동력선 사용 어가	1종겸업	강원, 경북
9	무동력선, 동력선 사용 어가	1종겸업	전북, 전남, 제주
10	무동력선, 동력선 사용 어가	1종겸업	부산, 울산, 경남
11	무동력선, 동력선 사용 어가	2종겸업	경기, 인천, 충남
12	무동력선, 동력선 사용 어가	2종겸업	강원, 경북
13	무동력선, 동력선 사용 어가	2종겸업	전북, 전남, 제주
14	무동력선, 동력선 사용 어가	2종겸업	부산, 울산, 경남
15	양식업	전업	전북, 전남, 제주
16	양식업	전업	전북, 전남, 제주 제외 나머지
17	양식업	1종겸업	전북, 전남, 제주
18	양식업	1종겸업	전북, 전남, 제주 제외 나머지
19	양식업	2종겸업	전북, 전남, 제주
20	양식업	2종겸업	전북, 전남, 제주 제외 나머지

<표 4.28> 어가경제조사의 무응답 대체군별 평균 및 표준편차(단위:만원) 방안 2

대체군	표본크기	어가소득		어업소득	
		평균	표준편차	평균	표준편차
1	31	1442.51	734.80	844.35	568.36
2	66	2116.27	1949.78	334.37	529.42
3	58	2189.66	1487.29	488.17	470.25
4	56	3183.20	2497.32	2735.50	2527.88
5	66	2590.70	1884.34	1915.34	1823.36
6	52	2081.11	1691.93	1802.65	1685.42
7	37	3432.58	3018.28	2488.86	2821.62
8	88	2572.91	1880.54	1748.14	1466.47
9	102	2678.66	1957.93	1684.29	1382.45
10	91	2499.08	1561.74	1598.27	1179.59
11	28	3353.88	2998.51	1011.54	1241.55
12	45	2606.08	1367.53	664.70	466.89
13	62	2589.31	1844.02	562.00	618.13
14	60	2696.19	1938.37	480.66	609.87
15	46	3404.65	3244.06	3099.83	3252.46
16	27	2662.39	2809.40	2259.68	2793.45
17	71	2659.89	2352.11	2014.38	2200.15
18	48	3257.21	2614.93	2309.48	2481.60
19	40	2254.60	1400.14	563.83	740.24
20	53	2378.79	1440.09	533.07	608.70

(3) 방안 3

방안 3은 방안 1과 마찬가지로 어가의 어업형태와 전·겸업 구분을 우선적으로 고려하여 무응답 대체군을 형성하는 방안이다. 다만, 동력선 사용어가와 양식업 어가에 대해서 선박톤수와 양식장 면적을 무응답 대체군 형성과정에서 이용하고, 이후에 지역을 대체군 세분 과정에서 이용하는 방안이다.

다음 <표 4.29>은 방안 3의 무응답 대체군 현황이다. 한편, <표 4.30>는 설정된 무응답 대체군별 어가소득과 어업소득 변수에 대해서 평균값과 표준편차를 구하여 정리한 것이다.

<표 4.29> 어가경제조사를 위한 무응답 대체군 구분 : 방안 3

대체군	어업형태	전겸업	구분 방법
1	비어선 어가	전업, 1종겸업	전체
2	비어선 어가	2종겸업	전북, 전남, 제주
3	비어선 어가	2종겸업	전북, 전남, 제주 제외 나머지
4	무동력선, 동력선 사용 어가	전업어가	선박 톤수<5.0
5	무동력선, 동력선 사용 어가	전업어가	선박 톤수>=5.0
6	무동력선, 동력선 사용 어가	1종겸업	경기, 인천, 충남
7	무동력선, 동력선 사용 어가	1종겸업	강원, 경북 & 선박 톤수 <2.5
8	무동력선, 동력선 사용 어가	1종겸업	강원, 경북 & 선박 톤수 >=2.5
9	무동력선, 동력선 사용 어가	1종겸업	전북, 전남, 제주 & 선박 톤수<2.5
10	무동력선, 동력선 사용 어가	1종겸업	전북, 전남, 제주 & 선박 톤수>=2.5
11	무동력선, 동력선 사용 어가	1종겸업	부산, 울산, 경남 & 선박 톤수<2.5
12	무동력선, 동력선 사용 어가	1종겸업	부산, 울산, 경남 & 선박 톤수>=2.5
13	무동력선, 동력선 사용 어가	2종겸업	경기, 인천, 충남
14	무동력선, 동력선 사용 어가	2종겸업	강원, 경북
15	무동력선, 동력선 사용 어가	2종겸업	전북, 전남, 제주
16	무동력선, 동력선 사용 어가	2종겸업	부산, 울산, 경남
17	양식업	전업	전북, 전남, 제주
18	양식업	전업	전북, 전남, 제주 제외 나머지
19	양식업	1종겸업	전북, 전남, 제주
20	양식업	1종겸업	전북, 전남, 제주 제외 나머지
21	양식업	2종겸업	양식장 면적<=0.8
22	양식업	2종겸업	양식장 면적>0.8

<표 4.30> 어가경제조사의 무응답 대체군별 평균 및 표준편차(단위:만원) 방안 3

대체군	표본크기	어가소득		어업소득	
		평균	표준편차	평균	표준편차
1	31	1442.61	734.80	844.38	568.36
2	66	2126.27	1949.78	334.37	529.42
3	58	2189.66	1487.29	488.17	470.25
4	111	2019.83	1406.39	1593.04	1272.41
5	63	3702.57	2606.87	3119.22	2752.81
6	37	3432.58	3018.28	2488.86	2821.62
7	52	2081.64	1048.67	1306.26	704.03
8	36	3282.51	2512.30	2386.42	1978.43
9	34	2068.56	997.05	1374.65	852.92
10	68	2983.71	2237.38	1839.11	1565.16
11	50	2039.81	1126.35	1311.81	921.59
12	41	3059.17	1829.89	1947.60	1364.80
13	28	3353.88	2998.51	1011.54	1241.55
14	45	2606.08	1367.53	664.70	466.89
15	62	2589.31	1844.02	562.00	618.13
16	60	2696.19	1938.37	480.66	609.87
17	46	3404.65	3244.06	3099.83	3252.46
18	27	2662.39	2809.40	2259.68	2793.45
19	71	2659.89	2352.11	2014.38	2200.15
20	48	3257.21	2614.93	2309.48	2481.6
21	46	1723.81	1098.89	362.61	483.09
22	47	2914.15	1454.88	726.08	767.34

(4) 방안 4

방안 4는 방안 3과 마찬가지로 어가의 어업형태와 전·겸업 구분을 우선적으로 고려하여 무응답 대체군을 형성하는 방안이다. 다만, 지역을 우선적으로 고려하여 대체군을 형성하고, 이후에 선박톤수와 양식장 면적을 무응답 대체군 세분하는 과정에서 이용하는 방안이다.

다음 <표 4.31>은 방안 4의 무응답 대체군 현황이다. 방안 4에서 대체군은 모두 23개로 구분된다. 한편, <표 4.32>은 설정된 무응답 대체군별 어가소득과 어업소득 변수에 대해서 평균값과 표준편차를 구하여 정리한 것이다.

<표 4.31> 어가경제조사를 위한 무응답 대체군 구분 : 방안 4

대체군	어업형태	전·겸업	구분 방법
1	비어선 어가	전업, 1종겸업	전체
2	비어선 어가	2종겸업	전북, 전남, 제주
3	비어선 어가	2종겸업	전북, 전남, 제주 제외 나머지
4	무동력선, 동력선 사용 어가	전업어가	경기, 인천, 충남, 강원, 경북
5	무동력선, 동력선 사용 어가	전업어가	전북, 전남, 제주
6	무동력선, 동력선 사용 어가	전업어가	부산, 울산, 경남
7	무동력선, 동력선 사용 어가	1종겸업	경기, 인천, 충남
8	무동력선, 동력선 사용 어가	1종겸업	강원, 경북 & 선박 톤수 <2.5
9	무동력선, 동력선 사용 어가	1종겸업	강원, 경북 & 선박 톤수 >=2.5
10	무동력선, 동력선 사용 어가	1종겸업	전북, 전남, 제주 & 선박 톤수<2.5
11	무동력선, 동력선 사용 어가	1종겸업	전북, 전남, 제주 & 선박 톤수>=2.5
12	무동력선, 동력선 사용 어가	1종겸업	부산, 울산, 경남 & 선박 톤수<2.5
13	무동력선, 동력선 사용 어가	1종겸업	부산, 울산, 경남 & 선박 톤수>=2.5
14	무동력선, 동력선 사용 어가	2종겸업	경기, 인천, 충남
15	무동력선, 동력선 사용 어가	2종겸업	강원, 경북
16	무동력선, 동력선 사용 어가	2종겸업	전북, 전남, 제주
17	무동력선, 동력선 사용 어가	2종겸업	부산, 울산, 경남
18	양식업	전업	전북, 전남, 제주
19	양식업	전업	전북, 전남, 제주 제외 나머지
20	양식업	1종겸업	전북, 전남, 제주
21	양식업	1종겸업	전북, 전남, 제주 제외 나머지
22	양식업	2종겸업	전북, 전남, 제주
23	양식업	2종겸업	전북, 전남, 제주 제외 나머지

<표 4.32> 어가경제조사의 무응답 대체군별 평균 및 표준편차(단위:만원) 방안 4

대체군	표본크기	어가소득		어업소득	
		평균	표준편차	평균	표준편차
1	31	1442.61	734.80	844.38	568.36
2	66	2126.27	1949.78	334.37	529.42
3	58	2189.66	1487.29	488.17	470.25
4	111	2019.83	1406.39	1593.04	1272.41
5	63	3702.57	2606.87	3119.22	2752.81
6	37	3432.58	3018.28	2488.86	2821.62
7	52	2081.64	1048.67	1306.26	704.03
8	36	3282.51	2512.30	2386.42	1978.43
9	34	2068.56	997.05	1374.65	852.92
10	68	2983.71	2237.38	1839.11	1565.16
11	50	2039.81	1126.35	1311.81	921.59
12	41	3059.17	1829.89	1947.60	1364.80
13	28	3353.88	2998.51	1011.54	1241.55
14	45	2606.08	1367.53	664.70	466.89
15	62	2589.31	1844.02	562.00	618.13
16	60	2696.19	1938.37	480.66	609.87
17	46	3404.65	3244.06	3099.83	3252.46
18	27	2662.39	2809.40	2259.68	2793.45
19	71	2659.89	2352.11	2014.38	2200.15
20	48	3257.21	2614.93	2309.48	2481.6
21	46	1723.81	1098.89	362.61	483.09
22	40	2254.60	1400.14	563.83	740.24
23	53	2378.79	1440.09	533.07	608.70

(5) 무응답 대체군 형성 방안의 비교

앞서 제시한 네 가지의 무응답 대체군 작성 방안을 농가경제조사의 무응답 대체군 구성에서 사용했던 비교방법에 따라서 살펴보자.

각 방안별 대체군 형성 효과는 중요한 변수에 대해서 다음 값을 계산하여 살펴볼 수 있다. 무응답 대체군 형성 효과는 백분율로 나타내어 그 값이 커질수록 무응답 대체군이 효과적으로 구성되었다는 것을 의미한다. 본 연구에서는 네 가지 방안을 비교하기 위하여 고려하는 주요변수는 어가소득, 어업소득, 겸업소득, 사업외 소득 등이다.

<표 4.33> 각 방안별 주요변수에 대한 무응답 대체군 형성 효과 비교

구 분	어가소득	어업소득	겸업소득	사업외 소득
방안 1	1.6%	16.4%	9.3%	17.0%
방안 2	2.3%	18.1%	14.4%	19.8%
방안 3	6.1%	20.9%	14.9%	21.7%
방안 4	3.7%	19.0%	14.6%	20.5%

분석결과를 보면 우선적으로 표본설계 상의 지역 구분 대신에 본 연구에서 제시한 지역 구분 방법을 이용하여 대체군을 형성하는 방안이 효율적임을 알 수 있다. 즉, 표본설계 상의 지역구분을 사용한 방안 1에 비해서 새로운 지역 구분 방법을 사용한 방안 2가 효율적이다.

또한 선박톤수와 어장면적을 대체군 형성 과정에서 고려하는 것이 효율적임을 알 수 있다. 이 사실은 이 두 변수를 대체군 형성 과정에서 이용한 방안 3과 4가 효과적이라는 사실에서 알 수 있다.

분석결과를 종합할 때 제시된 방안 중에서 방안 3이 가장 효과적인 것으로 나타났다. 방안 3에 따라서 대체군을 형성하게 되면 무응답 대체로 인한 변동을 어가소득 변수에 대해서는 약 6%, 어업소득 변수는 약 21%, 겸업소득 변수는 15%, 사업외 소득 변수는 약 22%를 낮출 수 있다. 본 연구에서는 무응답 대체군 형성방안으로 방안 3을 제안한다.

5. 가중핫덱 대체 및 분산추정

통상적으로 보조정보가 없는 경우는 핫덱 대체가 널리 쓰이며, 활용가능한 보조 정보가 있는 경우는 비대체나 회귀대체가 유용하게 쓰인다. 통상적인 문헌에 나오는 대체 방법은 대부분 단순임의표본을 전제로 했을 경우에 개발된 방법들이다. 그런데 농어가경제조사의 경우는 서로 다른 추출확률로 인하여 표본 가구들이 서로 다른 가중치를 가지고 있으므로 가중치를 사용하는 대체방법을 고찰해야 한다.

가중치가 있는 경우에 사용할 수 있는 핫덱 대체 방법은 가중핫덱 대체법이다. 이 방법은 Rao & Shao (1992)에 의해서 개발된 방법으로 대체군과 가중치를 이용하면 농어가경제조사에 적합한 방법이다.

5.1 가중핫덱대체

층화다단표본에 무응답이 발생한 경우를 고찰하자. 편의상 응답 메카니즘은 전 조사단위에서 같은 확률을 가지고 무응답이 발생한 경우를 고려하고, 대체 군은 1 개인 경우를 고려하자. 표본 중에서 응답을 한 집단을 s_r 로 표시하고 무응답이 발생한 집단을 s_m 으로 표시하자. 그리고 y_{hij}^* 를 대체된 데이터라고 하자. 그러면 대체 후 추정량은 다음과 같이 표현된다.

$$\bar{y}_I = \frac{\sum_{(hik) \in s_r} w_{hik} y_{hik} + \sum_{(hik) \in s_m} w_{hik} y_{hik}^*}{\sum_s w_{hik}} \quad (5.1)$$

위의 추정량의 성질은 대체 방법에 따라 다르다. 보조 변수가 없는 경우, 쉽게 사용할 수 있는 대체 방법으로는 평균대체, 핫덱 대체 등을 고려할 수 있는데 평균대체는 사용하기는 편리하지만 대체후 표본분포에 심한 왜곡이 발생하여 바람직하지 않은 것으로 알려져 있고, 핫덱 대체는 대체후 표본분포를 유지하는 장점이 있지만, 가중치가 있는 경우는 핫덱 대체 후 추정량이 편향이 되는 단점이 있다(Rao & Shao, 1992, p.816). 따라서 비편향성을 유지하면서 핫덱대체를 하려면 가중치를 이용하는 가중핫덱 대체를 해야 한다. 이 방법은 층화다단추출인 경우에 Rao & Shao (1992)에 의해서 제안 되었는데 그 절차는 다음과 같다.

가중햇덱 절차

- (i) 응답군 s_r 에서 무응답을 대체할 단위, (gil) ,를 가중치 $w_{gil} / \sum_{(hik) \in s_r} w_{hik}$ 에 비례하여 복원으로 뽑는다.
- (ii) 선정된 단위를 이용하여 무응답을 대체한다. $y_{hik}^* = y_{gil}$.

가중 햇덱 방법으로 무응답을 대체한 후 만든 대체후 추정량 \hat{Y}_I 는 모평균에 대한 근사비편향 추정량이 된다.

5.2 잭나이프 분산추정

일반적으로 무응답 대체를 할 때에 많이 나타나는 오류는 모총계 추정보다는 분산추정에서 많이 나타난다. 무응답을 대체하게 되면 대체로 인한 변동이 추가되어 대체 후 추정량 \bar{y}_I 의 분산은 무응답이 없는 추정량 \bar{y}_s 의 분산보다 크다.

$$Var\{\bar{y}_I\} = Var\{\bar{y}_s\} + V_{imp}$$

여기서 V_{imp} 는 대체로 인해 발생하는 분산이다. 그런데 대체 후 데이터를 마치 응답 데이터인 것처럼 취급하여 추정량을 만들고, 통상적인 분산추정량을 사용하게 되면 대체 후 분산을 과소 추정하게 된다. 대체 후 분산의 과소추정에 대한 논의는 Sarndal (1992)에 잘 나와 있다. 무응답률이 높을수록 과소 추정의 정도는 증가한다. 따라서 이러한 과소 추정의 문제를 해결하기 위해서는 분산추정량에 대체분산을 동시에 추정하는 항을 추가하여 분산을 비편향 추정할 수 있도록 해야 한다.

수정된 잭나이프 분산추정법을 알아보기 위하여 다음의 기호를 도입하자. g 층에서 j 번째 PSU를 제외한 통계량을 다음과 같이 약속하자.

$$\hat{S}(gj) = \sum_{(hik) \in s_r, h \neq g} w_{hik} y_{hik} + \frac{n_g}{n_g - 1} \sum_{(gik) \in s_r, i \neq j} w_{gik} y_{gik} \quad (5.2)$$

$$\hat{T}(gj) = \sum_{(hik) \in s_r, h \neq g} w_{hik} + \frac{n_g}{n_g - 1} \sum_{(gik) \in s_r, i \neq j} w_{gik} \quad (5.3)$$

그리고 다음을 정의하자. $(hik) \in s_m$ 인 (hik) 에 대하여,

$$Z_{hik}^{*(g)} = \begin{cases} y_{hik}^* + \frac{\widehat{\mathcal{S}}(g) - \widehat{\mathcal{T}}}{\widehat{\mathcal{T}}(g)} & (h) \neq (g) \\ y_{hik}^* & (h) = (g) \end{cases} \quad (5.4)$$

위의 식을 이용하여 (g) 번째 PSU를 제외한 후 만드는 잭나이프 반복 추정량은 다음과 같다.

$$\widehat{Y}_I^a(g) = \widehat{\mathcal{S}}(g) + \sum_{(hik) \in s_{np}, h \neq g} W_{hik} Z_{hik}^{*(g)} + \frac{n_g}{n_g - 1} \sum_{(hik) \in s_{np}, i \neq j} W_{gik} Z_{gik}^{*(g)} \quad (5.5)$$

마지막으로 수정된 잭나이프 분산추정량은 다음과 같다.

$$V_J(\widehat{Y}_I) = \sum_{g=1}^L \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} (\widehat{Y}^a(g) - \widehat{Y}_I)^2 \quad (5.6)$$

위 추정량은 대체 후 분산추정량 $Var\{\widehat{Y}_I$ 외 근사 일치추정량이다 (Rao & Shao, 1992).

위에 식 (5.6)은 모총계 추정량의 분산추정량이다. 따라서 모평균 추정량의 분산추정량을 구하기 위해서는 약간의 변형이 필요하다. 모평균을 추정하기 위해서는 모집단 크기 M 을 추정하는 과정이 선행되어야 한다. 모총계 추정과 동일한 방법을 이용하되 $y_{hik} = 1$ 을 치환하면 된다.

$$\widehat{\mathcal{S}}_M(g) = \sum_{(hik) \in s_r, h \neq g} W_{hik} + \frac{n_g}{n_g - 1} \sum_{(gik) \in s_r, i \neq j} W_{gik} \quad (5.7)$$

$$\widehat{\mathcal{T}}_M(g) = \sum_{(hik) \in s_r, h \neq g} W_{hik} + \frac{n_g}{n_g - 1} \sum_{(gik) \in s_r, i \neq j} W_{gik} \quad (5.8)$$

그리고 다음을 정의하자. $(hik) \in s_m$ 인 (hik) 에 대하여,

$$Z_{hik}^{*(g)} = \begin{cases} y_{hik}^* + \frac{\widehat{\mathcal{S}}_M(g) - \widehat{\mathcal{T}}_M}{\widehat{\mathcal{T}}_M(g)} & (h) \neq (g) \\ y_{hik}^* & (h) = (g) \end{cases} \quad (5.9)$$

= 1

위의 식을 이용하여 (g) 번째 PSU를 제외한 후 만드는 잭나이프 반복 추정량은 다음과 같다.

$$\widehat{M}_I^a(g) = \sum_{(hik) \in s, h \neq g} W_{hik} + \frac{n_g}{n_g - 1} \sum_{(hik) \in s, i \neq j} W_{gik} \quad (5.10)$$

그리고 모평균에 대한 잭나이프 반복 추정량은

$$\bar{y}_I^a(gj) = \hat{\Upsilon}_I^a(gj) / \hat{\mathcal{M}}_I^a(gj) \quad (5.11)$$

이며 수정된 잭나이프 분산추정량을 다음과 같다.

$$v_j(\bar{y}_I) = \sum_{g=1}^L \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} (\bar{y}^a(gj) - \bar{y}_I)^2 \quad (5.12)$$

5.3 모의실험

농어가경제조사의 주요 변수에 대하여 대체후 추정량의 비편향성과 잭나이프 분산추정량의 비편향성을 알아보기 위하여 모의실험을 실시하였다. 무응답 가구는 2, 4, 6, 8, 10개를 고려하였고 응답 메커니즘은 등확률 메커니즘을 고려하였다. 동일한 시행을 독립적으로 500번 반복하여 500개의 무응답을 포함한 표본을 생성하고, 각 각에서 가중합택 대체를 실시하였다. 그리고 500개의 대체후 추정치를 구하여 평균을 구하였다.

$$AVE = \frac{1}{K} \sum_{k=1}^K \bar{y}_I^{(k)}, \quad K = 500$$

여기서 k 번째 반복에서 구한 추정치는 다음과 같다.

$$\bar{y}_I = \frac{\sum_{(hik) \in s_r} w_{hik} y_{hik} + \sum_{(hik) \in s_m} w_{hik} y_{hik}^*}{\sum_{(hik) \in s} w_{hik}}$$

분산추정량으로는 통상적인 잭나이프 분산추정량(naive Jackknife variance estimator), v_{naive} 와 수정된 잭나이프 분산추정량 v_j 를 구하였다. 두 추정량의 효율은 대체 후 추정량 \bar{y}_I 의 분산과 비교하여 평가할 수 있다. 500개의 분산추정치들의 평균과 대체후 추정량의 분산을 비교하여 분산추정량의 편향과 안정성을 검토하였다.

- 상대편향 : $RB(\%) = \sum_{k=1}^K \frac{(v_k - V(\bar{y}_I))/K}{V(\bar{y}_I)} \times 100$
- 상대안정성 : $RS(\%) = \sum_{k=1}^K \frac{\sqrt{(v_k - V(\bar{y}_I))^2/K}}{V(\bar{y}_I)} \times 100$

상대편향이 음수이면 분산추정치가 분산을 과소 평가하고 있다는 의미이며, 양수이면 과대 평가하고 있다는 의미이다. 또한 상대안정성은 작을수록 참값에 가깝다는 뜻이 된다.

아래의 <표 5.1>-<표 5.4>은 농가경제조사의 주요변수 4개에 대한 모의실험 결과이고, <표 5.5>-<표 5.8>는 어가경제조사의 주요변수에 대한 결과이다. 데이터는 2002년 1월 데이터를 이용하였고, 계산의 편의를 위하여 단위를 10,000원으로 변환하였다.

<표 5.1> 잭나이프 분산추정량의 상대편향과 상대안정성 : 농가소득

무응답 수	AVE	Mean			상대편향(%)		상대안정성(%)	
2	190.76	119.37	119.67	119.43	-0.0444	0.2070	4.6845	5.7970
4	190.75	119.17	119.62	119.48	-0.2583	0.1156	6.5151	7.2131
6	190.72	119.06	119.66	119.56	-0.4172	0.0838	9.5855	9.5552
8	190.77	119.13	119.89	119.74	-0.5120	0.1198	14.0814	12.0143
10	190.75	118.87	119.77	119.74	-0.7280	0.0261	15.7698	13.0780

<표 5.2> 잭나이프 분산추정량의 상대편향과 상대안정성 : 농업소득

무응답 수	AVE	Mean			상대편향(%)		상대안정성(%)	
2	66.84	89.23	89.45	89.39	-0.1795	0.0742	4.9277	5.5075
4	66.87	89.15	89.49	89.47	-0.3644	0.0167	8.9158	7.7803
6	66.82	89.06	89.52	89.53	-0.5274	-0.0193	11.9367	9.8396
8	66.83	88.99	89.56	89.66	-0.7482	-0.1096	15.8874	12.1535
10	66.79	89.06	89.75	89.82	-0.8389	-0.0758	22.6150	16.3371

<표 5.3> 잭나이프 분산추정량의 상대편향과 상대안정성 : 겸업소득

무응답 수	AVE	Mean			상대편향(%)		상대안정성(%)	
2	125.33	123.10	123.45	123.21	-0.0937	0.1916	3.6601	5.3596
4	125.29	122.91	123.41	123.28	-0.2986	0.1096	5.9939	6.6899
6	125.32	123.00	123.66	123.43	-0.3523	0.1893	10.5666	9.6005
8	125.29	122.78	123.61	123.40	-0.5023	0.1633	10.9329	11.5393
10	125.30	123.51	123.79	122.80	-0.5714	0.2241	13.8215	13.1917

<표 5.4> 잭나이프 분산추정량의 상대편향과 상대안정성 : 사업외소득

무응답 수	AVE	Mean			상대편향(%)		상대안정성(%)	
2	58.40	23.13	23.21	23.10	-0.1142	0.3541	2.8836	7.8495
4	58.39	23.08	23.22	23.15	-0.3039	0.2883	7.8540	10.0477
6	58.40	23.06	23.23	23.17	-0.4560	0.2680	11.3682	12.9566
8	58.40	23.03	23.22	23.19	-0.6958	0.1547	14.2912	13.8275
10	58.39	23.04	23.27	23.21	-0.7296	0.2366	17.9616	15.8575

<표 5.5> 잭나이프 분산추정량의 상대편향과 상대안정성 : 여가소득

무응답 수	AVE	평균			상대편향(%)		상대안정성(%)	
2	193.56	179.32	179.94	179.61	-0.1629	0.1823	23.3670	18.3966
4	193.52	178.41	179.62	179.50	-0.6122	0.0665	19.1682	13.9881
6	193.54	178.44	180.24	179.94	-0.8330	0.1679	26.5608	20.0207
8	193.51	178.43	180.85	181.10	-1.4743	-0.1403	46.7094	28.4109
10	193.48	177.96	180.99	181.19	-1.7845	0.1119	48.5796	31.7441

<표 5.6> 잭나이프 분산추정량의 상대편향과 상대안정성 : 어업소득

무응답 수	AVE	평균			상대편향(%)		상대안정성(%)	
2	105.24	138.84	139.32	139.37	-0.3744	-0.0329	11.7437	9.2362
4	105.23	138.83	139.79	139.98	-0.8215	-0.1386	30.1480	20.9044
6	105.19	138.55	139.99	140.01	-1.0467	-0.0130	30.5357	21.9496
8	105.19	138.64	140.55	140.63	-1.4161	-0.0556	49.0301	32.2641
10	105.17	138.78	141.16	141.02	-1.5870	0.1011	61.0757	39.5155

<표 5.7> 잭나이프 분산추정량의 상대편향과 상대안정성 : 겸업소득

무응답 수	AVE	평균			상대편향(%)		상대안정성(%)	
2	25.65	27.30	27.40	27.42	-0.4354	-0.0890	11.4185	9.1993
4	25.68	27.32	27.51	27.51	-0.6888	-0.0006	23.1636	16.2131
6	25.69	27.33	27.62	27.57	-0.8646	0.1698	34.0905	25.6104
8	25.68	27.26	27.64	27.64	-1.3676	-0.0045	41.5188	30.3492
10	25.70	27.78	27.78	27.83	-1.8847	-0.1839	64.6242	41.2877

<표 5.8> 잭나이프 분산추정량의 상대편향과 상대안정성 : 사업외소득

무응답 수	AVE	평균			상대편향(%)		상대안정성(%)	
2	33.72	12.05	12.09	12.07	-0.1891	0.1499	7.6397	6.8583
4	33.73	12.05	12.13	12.13	-0.6799	-0.0021	25.3509	16.9296
6	33.71	12.05	12.17	12.19	-1.1017	-0.0944	38.3484	23.5835
8	33.74	12.25	12.23	12.07	-1.4868	-0.1578	52.3058	32.8415
10	33.75	12.06	12.27	12.31	-2.0246	-0.3732	65.6279	40.4876

농가경제조사 데이터와 어가경제조사 데이터를 이용하여 모의실험을 한 결과 다음과 같은 사항을 알 수 있다.

첫째, 통상적인 분산추정량(v_{naive})은 무응답 가구수가 커지면 상대편향이 증가하지만, 잭나이프 분산추정량(v_j)는 상대편향이 증가하는 경향이 없다. 즉, 잭나이프 분산추정량이 근사적으로 비편향성을 갖는다는 점이 경험적으로 나타나고 있다고 볼 수 있다. 상대편향의 크기는 잭나이프 분산추정량이 통상적인 분산추정량보다 대체로 작다.

둘째, 무응답 가구수가 증가하면 상대안정성은 수치가 증가한다. 그러나 잭나이프 분산추정량의 수치가 통상적인 분산추정량의 수치보다 대체로 작게 나타난다. 이와 같은 두 가지 사실을 종합하면 잭나이프 분산추정량이 통상적인 분산추정량보다 우수하며, 농어가경제조사에 적합하다는 사실을 알 수 있다.

5.4 잭나이프 분산 추정 : 다중 대체군

표본 전체에서 무응답을 대체하는 것 보다는 적절한 대체군을 만든 뒤, 각 대체군에서 무응답을 대체하는 것이 더 효율적이다. 농어가경제조사에서 대체군을 만드는 방법은 앞의 4장에서 설명한 바 있다. 이번 절에서는 각 대체군에서 무응답을 대체한 후 잭나이프 분산추정법을 소개하기로 한다.

절차 1 : 대체군 형성

절차 2 : 대체군 v 에서 가중하택 실시

(i) 응답군 s_{IV} 에서 무응답을 대체할 USU, (gl) ,를 가중치

$$w_{gl} / \sum_{(hik) \in s_{IV}} w_{hik} \text{에 비례하여 복원으로 뽑는다.}$$

(ii) 선정된 USU를 이용하여 무응답을 대체한다. $y_{hik}^* = y_{gl}$.

절차 3 : 대체 후 추정량 계산 : 대체군 v 에서

$$\hat{Y}_{IV} = \sum_{(hik) \in A_{IV}} w_{hik} y_{hik} + \sum_{(hik) \in A_{mv}} w_{hik} y_{hik}^* \quad (5.13)$$

$$\hat{M}_{IV} = \sum_{(hik) \in A_{IV}} w_{hik} + \sum_{(hik) \in A_{mv}} w_{hik} = \sum_{(hik) \in S_v} w_{hik} \quad (5.14)$$

을 만든 후, 대체군을 합하여 다음을 구한다.

$$\hat{Y}_I = \sum_v \hat{Y}_{IV}, \quad \hat{M}_I = \sum_v \hat{M}_{IV} = \sum_{(hik) \in S} w_{hik}$$

$$\bar{y}_I = \hat{Y}_I / \hat{M}_I$$

절차 4 : 잭나이프 반복추정치 계산. 대체군 v 에서 다음을 계산한다. 대체군 v 에서 g 층의 j 번째 PSU를 제외한 통계량을 다음과 같이 약속하자.

$$\hat{\mathcal{S}}(gj(v)) = \sum_{(hik) \in S_{r(v)}, h \neq g} w_{hik} y_{hik} + \frac{n_g}{n_g - 1} \sum_{(gik) \in S_{r(v)}, i \neq j} w_{gik} y_{gik} \quad (5.15)$$

$$\hat{\mathcal{T}}(gj(v)) = \sum_{(hik) \in S_{r(v)}, h \neq g} w_{hik} + \frac{n_g}{n_g - 1} \sum_{(gik) \in S_{r(v)}, i \neq j} w_{gik} \quad (5.16)$$

그리고 다음을 정의하자. $(hik) \in S_{m(v)}$ 인 (hik) 에 대하여,

$$Z_{hik}^{*(g)} = \begin{cases} y_{hik}^* + \frac{\hat{\mathcal{S}}(gj(v))}{\hat{\mathcal{T}}(gj(v))} - \frac{\hat{\mathcal{S}}(v)}{\hat{\mathcal{T}}(v)} & (h) \neq (g) \\ y_{hik}^* & (h) = (g) \end{cases} \quad (5.17)$$

위의 식을 이용하여 대체군 v 에서 (g) 번째 PSU를 제외한 후 만드는 잭나이프 반복 추정량은 다음과 같다.

$$\hat{Y}_{IV}^a(gj) = \hat{\mathcal{S}}(gj(v)) + \sum_{(hik) \in S_{m(v)}, h \neq g} w_{hik} Z_{hik}^{*(g)} + \frac{n_g(v)}{n_g(v) - 1} \sum_{(hik) \in S_{m(v)}, i \neq j} w_{gik} Z_{gik}^{*(g)} \quad (5.18)$$

$$\hat{M}_{IV}^a(gj) = \hat{\mathcal{T}}(gj(v)) + \sum_{(hik) \in S_{m(v)}, h \neq g} w_{hik} + \frac{n_g(v)}{n_g(v) - 1} \sum_{(hik) \in S_{m(v)}, i \neq j} w_{gik} \quad (5.19)$$

이제 각 대체군에서 만든 복제 추정량을 합하여 잭나이프 복제 추정량을 만든다.

$$\widehat{Y}_I^a(gj) = \sum_v \widehat{Y}_{N^a}^a(gj),$$

$$\widehat{M}_I^a(gj) = \sum_v \widehat{M}_{N^a}^a(gj)$$

$$\overline{y}_I^a(gj) = \widehat{Y}_I^a(gj) / \widehat{M}_I^a(gj)$$

절차 5 마지막으로 수장된 잭나이프 추정량은 다음과 같이 얻는다.

$$v_j(\overline{y}_I) = \sum_{g=1}^L \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} (\overline{y}_I^a(gj) - \overline{y}_I)^2 \quad (5.20)$$

잭나이프 분산추정을 하는 계산알고리즘을 SAS로 구현한 프로그램이 부록에 실려있다.

6. 결론

본 보고서는 농어가경제조사에서 발생하는 표본교체와 무응답 처리기법에 관한 연구 결과이다. 연구결과를 요약하면 다음과 같다.

일계부 무응답 처리

일계부에서 발생하는 무응답은 표본대체를 한다. 표본대체를 하기 위해서는 무응답 대체군을 만드는 일이 선행되어야 하는데, 전년도 연간 표본자료를 이용하여 연초에 만든다. 혹은 필요에 따라 최근자료를 이용하여 대체하고자 하는 달에서 가까운 시점에 만들어도 된다. 대체방법은 가중햇택 방법을 사용하며, 모평균은 대체후 가중평균을 사용한다. 대체후 가중평균에 대한 표준오차는 잭나이프 분산추정치를 이용하여 구한다.

교체가구 처리

표본 농어가에서 교체 사유가 발생하면 조사의 공백이 생기지 않도록 즉시 교체한다. 일계부는 교체시점부터 추가 작성하며, 원부는 연초수치의 소급 조사가 가능한 항목은 소급 조사하여 사용하고 소급조사가 어려운 항목은 (i) 교체시점을 연초로 간주하여 조사하거나 (ii) 연초 평가액 및 증감액을 추정한다. 증감액의 크기, 교체 시점, 현지 표본가구의 상황을 고려하여 정한다.

연간 모평균 추정

1월부터 12월까지의 조사가 완료되면 이듬해 1월에 일계부 데이터와 원부 데이터를 결합하여 주요변수에 대하여 연평균 추정치를 계산한다. 이 때, 대체군은 최근의 자료를 이용하여 만드는 것이 유리하다. 대체군을 이용하여 가중햇택 대체를 한 후, 일계부에 대한 완비 데이터를 만든다. 그리고 원부 데이터와 결합하여 연평균 추정치를 계산한다. 그리고 연평균 추정치에 대한 표준오차는 표본설계가 제안하는 분

산추정량을 이용하여 구한다.

연평균 추정치는 무응답 대체 및 표본 교체가 포함되었기 때문에 엄밀한 의미에서 표본설계가 제안하는 분산추정치가 모평균 추정치의 정확한 분산추정치는 아니다. 그렇지만 농어가경제조사에서는 무응답률이 높지 않아서 분산의 과소추정문제는 심각하지 않을 것으로 생각된다. 또한 월별로 대체가 실시되고, 연간 데이터는 월별 데이터의 합이기 때문에 연간 데이터에서 무응답 대체효과를 분리하여 이론적인 분산추정량을 구하는 일은 쉽지 않으며, 더 많은 연구를 필요로 한다. 향후 연구 과제로 남겨둔다.

참고문헌

- [1] 통계청(2003.7). 농가경제조사 농산물생산비조사 지침서
- [2] 통계청(2003.8). 어가경제조사 지침서.
- [3] 통계청(2003.6). 2002 농가경제통계.
- [4] 통계청(2003.6). 2002년 어가경제통계
- [5] 한국통계학회 (2002). 농가경제조사 표본설계. 연구보고서.
- [6] 한국통계학회 (2002). 어가경제조사 표본설계. 연구보고서.
- [7] Kovar, J.G. and Chen, E. (1994). Jackknife variance estimation of imputed survey data. *Survey Methodology*, 20, 45-52.
- [8] Krewski, D. and Rao, J.N.K. (1981). Inference from stratified samples : properties of the linearization, jackknife and balanced repeated replication methods. *Annals of Statistics*, 9, 1010-1019.
- [9] Rao, J.N.K. and Wu, C.F.J. and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18, 209-217.
- [10] Sarndal, C.E. (1992). Methods for estimating the precision of survey estimates when imputation has been used. *Survey Methodology*, 18, 241-252.
- [11] Yung, W. and Rao, J.N.K. (1996). Jackknife linearization variance estimators under stratified multi-stage sampling. *Survey Methodology*, 22, 23-31.

<부록> 잭나이프 분산추정량 계산 알고리즘(SAS 프로그램)

다음은 5장에서 설명한 잭나이프 분산추정을 구현하는 SAS 프로그램이다. 대체군은 3개를 고려하였고, 대체군 1과 2에서 무응답이 하나씩 발생하고 세 번째 대체군에서는 무응답이 발생하지 않은 경우는 고려하였다. 대체군의 수와 무응답 수가 증가한 경우는 약간의 변형을 통하여 사용가능하다.

```

/*-----
Jackknife Variance Estimation after Weighted Hot-Deck Imputation
  - multiple imputation class
  - sample size : 30
  - PSU : 10
  - imputation class : 3
Date : 2004. 10. 23
-----*/;

/*----- Original Data Set -----*/;

data pop;
  input num st psu ssu nh y1 y2 wt res ic ;
cards;
  1 1 1 1 3 20 1899 8.1111 1 1
  2 1 1 2 3 53 5931 8.1111 1 2
  3 1 1 3 3 48 4968 8.1111 1 2
  4 1 2 1 3 21 7082 8.1111 1 2
  5 1 2 2 3 13 1830 8.1111 1 1
  6 1 2 3 3 25 2628 8.1111 1 2
  7 1 3 1 3 18 1645 8.1111 1 1
  8 1 3 2 3 12 1408 8.1111 1 2
  9 1 3 3 3 35 3635 8.1111 1 2
  10 2 1 1 3 13 1457 7.7778 1 2
  11 2 1 2 3 17 1733 7.7778 1 2
  12 2 1 3 3 13 2042 7.7778 1 2
  13 2 2 1 3 7 779 7.7778 0 1
  14 2 2 2 3 70 6317 7.7778 1 3
  15 2 2 3 3 31 3073 7.7778 1 2
  16 2 3 1 3 13 1276 7.7778 1 1
  17 2 3 2 3 11 1201 7.7778 1 1
  18 2 3 3 3 35 3096 7.7778 0 2
  19 3 1 1 4 13 2092 11.7500 1 1
  20 3 1 2 4 11 1135 11.7500 1 1
  21 3 1 3 4 12 1736 11.7500 1 1
  22 3 2 1 4 11 1224 11.7500 1 1
  23 3 2 2 4 49 4798 11.7500 1 3
  24 3 2 3 4 13 1101 11.7500 1 1
  25 3 3 1 4 74 6635 11.7500 1 3

```

26	3	3	2	4	17	1868	11.7500	1	2
27	3	3	3	4	18	1719	11.7500	1	2
28	3	4	1	4	13	1989	11.7500	1	1
29	3	4	2	4	18	2616	11.7500	1	2
30	3	4	3	4	26	2067	11.7500	1	3

```

;

proc sort; by st psu ;

proc freq data=pop noprint;
  tables psu/ out=psul;
  by st ;

data psul;
  set psul;
  keep st psu;

proc freq data=psul noprint;
  tables st/out=psu2;

data psu;
  merge psul psu2;
  by st;
  nh=count;
  keep st psu nh;
  run;

proc means data=pop n noprint;
  var ssu;
  output out=mm n=mm;

data mm;
  set mm; keep mm;

proc means data=psu n noprint;
  var nh;
  output out=nn n=nn;
data nn;
  set nn; keep nn;
run;

data count;
  merge nn mm;
run;

%macro weighted_HD(y=,seed11=,seed12=, seed21=, seed22=, seed31=, seed32=, m1_size=,
m2_size=, m3_size=);

data pop1 pop2 pop3 ;

```

```

set pop;
    if ic=1 then output pop1;
else if ic=2 then output pop2;
else          output pop3;

/* ----- weighted HD Imputation : class 1 -----*/;

data response missing ;
    set pop1;
    if res=1 then output response;
else          output missing;

proc surveysselect data=response method=pps_sys n=&m1_size seed=&seed12 out=imput;
    size wt;

data imput;
    set imput;
    yi_star=&y;
    keep yi_star;

data miss;
    merge missing imput;

data response;
    set response;
    yi_star = &y ;

data completel;
    set response miss;
    yi=&y;

proc sort data=completel;
    by ic st psu ssu;

/* ----- weighted HD Imputation : class 2 -----*/;

data response missing ;
    set pop2;
    if res=1 then output response;
else          output missing;

proc surveysselect data=response method=pps_sys n=&m2_size seed=&seed22 out=imput;
    size wt;

data imput;
    set imput;
    yi_star=&y;
    keep yi_star;

```

```

data miss;
  merge missing imput;

data response;
  set response;
  yi_star = &y ;

data complete2;
  set response miss;
  yi=&y;

proc sort data=complete2;
  by ic st psu ssu;

/*----- class 3 -----*/;

data complete3;
  set pop3;
  res=1;
  yi=&y;
  yi_star=&y;
run;

proc sort data=complete3;
  by ic st psu ssu;
run;

/*----- complete data -----*/;
data complete;
  set completel complete2 complete3;
  run;

/*----- naive estiamtor -----*/;

proc surveymeans data=complete nobobs mean stderr var cv;
  stratum st ;
  cluster psu;
  var yi yi_star ;
  weight wt;
  ods output statistics=mystat;
run;

data mystat;
  set mystat;
  mean_y_s = yi_mean;
  var_y_s = yi_var;
  mean_y_i = yi_star_mean;
  v_naive = yi_star_var;
  keep mean_y_s var_y_s mean_y_i v_naive;

```



```

run;

/*----- input data -----*/;

proc iml;
use complete;
read all var {          st  psu  ssu  nh  wt  ic  res  yi  yi_star} into sample;
mattrib sample colname={ st  psu  ssu  nh  wt  ic  res  yi  yi_star};

use psu;
read all var {st psu nh} into index;
mattrib index colname={st psu nh};

use count;
read all var { mm  nn} into count;
mattrib count colname={ mm nn};

m=count[1,1];
n=count[1,2];

/*----- Estimator after imputation -----*/;

ww=0; wy=0;

do k=1 to m;
  do; ww=ww+sample[k,5];
      wy=wy+sample[k,5]*sample[k,9];
  end; end;

bar_y_i = wy/ww;

/*----- making matrix of weight -----*/;

weight_gj=j(m,n,0);

do gj=1 to n;
  do k=1 to m; do;
    if ((index[gj,1]=sample[k,1]) & (index[gj,2] =sample[k,2])) then
weight_gj[k,gj]=0;
    else if ((index[gj,1]=sample[k,1]) & (index[gj,2] ^=sample[k,2])) then

weight_gj[k,gj]=sample[k,4]/(sample[k,4]-1)*sample[k,5];
    else
weight_gj[k,gj]=sample[k,5];

  end; end; end;

mattrib weight_gj colname={ w11 w12 w13 w21 w22 w23 w31 w32 w33 w34 };

/*----- making matrix of stat_gj -----*/;

```

```

stat_gj =j(8,n,0);

do gj=1 to n;
  hat_s1 =0;   hat_t1 =0;   hat_s1_gj=0;   hat_t1_gj=0;
  hat_s2 =0;   hat_t2 =0;   hat_s2_gj=0;   hat_t2_gj=0;

do k=1 to m;
  if sample[k,6]=1 then do;
    hat_s1 = hat_s1 + sample[k,5]*sample[k,9]*sample[k,7];
    hat_t1 = hat_t1 + sample[k,5]*sample[k,7];
    hat_s1_gj= hat_s1_gj+ weight_gj[k,gj]*sample[k,9]*sample[k,7];
    hat_t1_gj= hat_t1_gj+ weight_gj[k,gj]*sample[k,7];
  end;

  else if sample[k,6]=2 then do;
    hat_s2 = hat_s2 + sample[k,5]*sample[k,9]*sample[k,7];
    hat_t2 = hat_t2 + sample[k,5]*sample[k,7];
    hat_s2_gj= hat_s2_gj+ weight_gj[k,gj]*sample[k,9]*sample[k,7];
    hat_t2_gj= hat_t2_gj+ weight_gj[k,gj]*sample[k,7];
  end;end;

  stat_gj[1,gj]=hat_s1_gj;   stat_gj[2,gj]=hat_t1_gj;
  stat_gj[3,gj]=hat_s1;     stat_gj[4,gj]=hat_t1;

  stat_gj[5,gj]=hat_s2_gj;   stat_gj[6,gj]=hat_t2_gj;
  stat_gj[7,gj]=hat_s2;     stat_gj[8,gj]=hat_t2;
end;

/*----- making matrix of z_values -----*/;

yz_gj =j(m,n,0);

do gj=1 to n;
  do k=1 to m;
    if sample[k,7]=1 then yz_gj[k,gj]=sample[k,9];
    else if sample[k,6]=1 then
yz_gj[k,gj]=sample[k,9]+(stat_gj[1,gj]/stat_gj[2,gj]-stat_gj[3,gj]/stat_gj[4,gj]);
    else if sample[k,6]=2 then
yz_gj[k,gj]=sample[k,9]+(stat_gj[5,gj]/stat_gj[6,gj]-stat_gj[7,gj]/stat_gj[8,gj]);
    else yz_gj[k,gj]=sample[k,9];
  end; end;

/*----- making replicates -----*/;

repli=j(3,n,0);

do gj=1 to n;
  hat_m_I =0;   hat_y_I =0;   hat_y_0 =0;

```

```

do k=1 to m;
  hat_m_I = hat_m_I + weight_gj[k,gj];
  hat_y_I = hat_y_I + weight_gj[k,gj]*yz_gj[k,gj];
  hat_y_0 = hat_y_0 + weight_gj[k,gj]*sample[k,9];
end;

repli[1,gj]=hat_m_i;  repli[2,gj]=hat_y_i/hat_m_i;  repli[3,gj]=hat_y_0/hat_m_i;

end;

/* ----- making statistics -----*/;

v_j = 0; v_n = 0;

do gj=1 to n;
  v_j = v_j + (index[gj,3]-1)/ index[gj,3] * (repli[2,gj]-bar_y_i)**2;
  v_n = v_n + (index[gj,3]-1)/ index[gj,3] * (repli[3,gj]-bar_y_i)**2;
end;

stat=j(1,3,0);
stat[1,1]= bar_y_i; stat[1,2]= v_j ; stat[1,3]= v_n;

vname={'mean_y_i', 'v_J', 'v_naive'};
create stat from stat[colname=vname];
append from stat;
close stat;

proc print data=stat;
quit;
%mend weighted_HD;

%weighted_HD(y=y1,seed11=1000,seed12=2000, seed21=3000, seed22=4000,
             seed31=5000, seed32=6000, m1_size=1, m2_size=1, m3_size=0 );
%weighted_HD(y=y2,seed11=1100,seed12=2200, seed21=3300, seed22=4400,
             seed31=5500, seed32=6600, m1_size=1, m2_size=1, m3_size=0 );

```

프로그램 수행 결과

Obs	mean_y_i	v_J	v_naive
1	23.8248	13.7872	13.1981

Obs	mean_y_i	v_J	v_naive
1	2654.94	95994.63	92289.80